

1W-47
334 477

NASA Contractor Report CR-1998-207910

An Extended Objective Evaluation of the 29-km Eta Model For Weather Support to the United States Space Program

Prepared By:
Applied Meteorology Unit

Prepared for:
Kennedy Space Center
Under Contract NAS10-96018

NASA
National Aeronautics and
Space Administration

Office of Management

Scientific and Technical
Information Program

1998

Attributes and Acknowledgments

NASA/KSC POC:
Dr. Francis J. Merceret
AA-C-1

Applied Meteorology Unit (AMU)

Paul Nutter
John Manobianco

Table of Contents

List of Figures.....	vi
List of Tables.....	viii
Executive Summary.....	ix
1.0 Introduction.....	1
1.1 Motivation for Objective Point Forecast Verification.....	1
1.2 Precedent and Progress	1
1.3 Applied Meteorology Unit Tasking.....	1
1.4 Report Objectives and Organization	2
2.0 Eta Model Overview.....	3
3.0 Data and Analysis Method	3
3.1 Data Collection Periods.....	3
3.2 Data Collection Methods.....	3
3.3 Statistical Analysis Methods	4
4.0 Surface Forecast Accuracy	4
4.1 Overall Summary and Interpretation	4
4.2 1996 Surface Forecast Accuracy	7
4.2.1 1996 Warm Season	7
4.2.1.1 Mean Sea-Level Pressure.....	7
4.2.1.2 2-m Temperature.....	7
4.2.1.3 2-m Dew Point Temperature.....	7
4.2.1.4 10-m Wind Speed	9
4.2.1.5 10-m Wind Direction	9
4.2.2 1996 Cool Season	9
4.2.2.1 Mean Sea-Level Pressure.....	9
4.2.2.2 2-m Temperature.....	9
4.2.2.3 2-m Dew Point Temperature.....	9
4.2.2.4 10-m Wind Speed	11

4.2.2.5	10-m Wind Direction	11
4.3	Changes in Surface Forecast Accuracy During 1997	11
4.3.1	1997 Warm Season	12
4.3.1.1	Mean Sea-Level Pressure	12
4.3.1.2	2-m Temperature.....	12
4.3.1.3	2-m Dew Point Temperature.....	12
4.3.1.4	10-m Wind Speed	15
4.3.1.5	10-m Wind Direction	15
4.3.1.6	Summary of 1997 Warm Season Changes	15
4.3.2	1997 Cool Season	16
4.3.2.1	Mean Sea-Level Pressure	16
4.3.2.2	2-m Temperature.....	16
4.3.2.3	2-m Dew Point Temperature	16
4.3.2.4	10-m Wind Speed	16
4.3.2.5	10-m Wind Direction	19
4.3.2.6	Summary of 1997 Cool Season Changes	19
4.4	10-m Wind Persistence.....	19
5.0	Upper Air Forecast Accuracy	20
5.1	Summary of Upper-Air Forecast Accuracy	20
5.2	Detailed Results.....	21
5.2.1	Temperature	21
5.2.2	Mixing Ratio	22
5.2.3	Wind Speed.....	22
5.2.4	Wind Direction.....	26
5.2.5	Geopotential Height	26
5.3	Forecast Error Growth.....	29
6.0	Convective Indices	31
6.1	Summary of Convective Index Error Characteristics	31

6.2 Convective Index Error Characteristics.....	32
6.2.1 Precipitable Water.....	32
6.2.2 Lifted Index.....	32
6.2.3 K-Index	36
6.2.4 Lifted Condensation Level.....	36
6.2.5 Convective Available Potential Energy.....	36
6.2.6 Convective Inhibition.....	40
6.2.7 Helicity.....	40
6.2.8 Microburst Day Potential Index	40
7.0 850 to 500-mb Layer-Averages	44
8.0 Wind Regime Stratification	44
9.0 Summary of Subjective Forecast Value.....	45
10.0 Summary and Lessons Learned	46
10.1 Surface Results	46
10.2 Upper-air Results.....	46
10.3 Convective Index Results	47
10.4 Layer Averages and Wind Regime Stratification	47
10.5 Lessons and Recommendations	47
11.0 References	48
Appendix A	51
Appendix B.....	52
Appendix C.....	54
Appendix D	58
Appendix E.....	59

List of Figures

Figure 4.1.	Bias (forecast – observed), RMS error, and error standard deviation for surface parameter forecasts from the 0300 UTC Meso-Eta cycle during the 1996 warm season.	8
Figure 4.2.	Bias (forecast – observed), RMS error, and error standard deviation for surface parameter forecasts from the 0300 UTC Meso-Eta cycle during the 1996 cool season	10
Figure 4.3.	1997 warm season bias (forecast – observed), annual difference of absolute bias (AB97 – AB96), and standardized Z statistics for the 0300 UTC Meso-Eta cycle.	13
Figure 4.4.	Comparison of annual changes (1997 - 1996) in mean forecasts and observations during the warm season.	14
Figure 4.5.	1997 cool season bias (forecast – observed), annual difference of absolute bias (AB97 – AB96), and standardized Z statistics for the 0300 UTC Meso-Eta cycle.	17
Figure 4.6.	Comparison of annual changes (1997 - 1996) in mean forecasts and observations during the cool season.	18
Figure 5.1.	Bias (forecast – observed), RMS error, and error standard deviation (°C) of temperature forecasts plotted as a function of pressure level	23
Figure 5.2.	Bias (forecast – observed), RMS error, and error standard deviation (g kg ⁻¹) of mixing ratio forecasts plotted as a function of pressure level	24
Figure 5.3.	Bias (forecast – observed), RMS error, and error standard deviation (m s ⁻¹) of wind speed forecasts plotted as a function of pressure level	25
Figure 5.4.	Bias (forecast – observed), RMS error, and error standard deviation (°) of wind direction forecasts plotted as a function of pressure level	27
Figure 5.5.	Bias (forecast – observed), RMS error, and error standard deviation (m) of geopotential height forecasts plotted as a function of pressure level	28
Figure 5.6.	Paired Z statistic plotted as a function of pressure level.....	30
Figure 6.1.	Quartile diagram showing the conditional distribution of observed PWAT (mm) associated with given forecast values and a histogram of the marginal distribution of forecasts.....	34
Figure 6.2.	Quartile diagram showing the conditional distribution of observed lifted index (°C) associated with given forecast values and a histogram of the marginal distribution of forecasts.....	35
Figure 6.3.	Quartile diagram showing the conditional distribution of observed K-index (°C) associated with given forecast values and a histogram of the marginal distribution of forecasts.....	37
Figure 6.4.	Quartile diagram showing the conditional distribution of observed LCL (mb) associated with given forecast values and a histogram of the marginal distribution of forecasts.....	38
Figure 6.5.	Quartile diagram showing the conditional distribution of observed CAPE (J kg ⁻¹) associated with given forecast values and a histogram of the marginal distribution of forecasts.....	39

Figure 6.6.	Quartile diagram showing the conditional distribution of observed CIN (J kg^{-1}) associated with given forecast values and a histogram of the marginal distribution of forecasts.....	41
Figure 6.7.	Quartile diagram showing the conditional distribution of observed helicity ($\text{m}^2 \text{s}^{-2}$) associated with given forecast values and a histogram of the marginal distribution of forecasts.	42
Figure 6.8.	Quartile diagram showing the conditional distribution of observed MDPI associated with given forecast values and a histogram of the marginal distribution of forecasts.	43

List of Tables

Table 1.1.	Meso-Eta verification parameters.....	2
Table 3.1.	Definition of seasonal verification periods and notable Eta model updates.	3
Table 4.1.	Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at XMR.....	5
Table 4.2.	Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at TBW.....	6
Table 4.3.	Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at EDW.....	6
Table 5.1.	Summary of Meso-Eta upper-air forecast error characteristics at XMR and TBW.....	20
Table 5.2.	Summary of Meso-Eta upper-air forecast error characteristics at EDW.....	21
Table 6.1.	Summary of Meso-Eta point forecast error characteristics for convective indices at XMR during the warm season (May-Aug).	32
Table 6.2.	Bias, RMS error and error standard deviation for warm season (May - August) convective parameters at XMR and TBW.....	33
Table 6.3.	Bias, RMS error and error standard deviation for cool season (October - January) convective parameters at XMR and TBW.....	33
Table 7.1.	Warm and cool season bias, RMS error, and error standard deviation for 1000- to 850-mb thickness and 850- to 500-mb layer-averaged relative humidity, wind speed, and wind direction.....	44
Table A.1.	Eta model attributes from Black (1994), Janjic (1994), and Rogers et al. (1996).	51
Table C.1.	Warm season biases for wind variables at XMR and EDW as a function of geopotential height.....	55
Table C.2.	Cool season biases for wind variables at XMR and EDW as a function of geopotential height.....	55
Table C.3.	Warm season RMS errors for wind variables at XMR and EDW as a function of geopotential height.	56
Table C.4.	Cool season RMS errors for wind variables at XMR and EDW a function of geopotential height.....	56
Table C.5.	Warm season error standard deviations for wind variables at XMR and EDW as a function of geopotential height.	57
Table C.6.	Cool season error standard deviations for wind variables at EDW as a function of geopotential height.	57

Executive Summary

This report describes the Applied Meteorology Unit's (AMU) objective verification of the National Centers for Environmental Prediction (NCEP) 29-km Eta (Meso-Eta) numerical weather prediction model. The verification was designed to identify the model's error characteristics for surface and upper-air point forecasts at Cape Canaveral Air Station (XMR), FL, Tampa Bay (TBW), FL, and Edwards Air Force Base (EDW), CA. These stations are selected because they are important for 45th Weather Squadron (45WS), Spaceflight Meteorology Group (SMG), and National Weather Service (NWS) Melbourne (MLB) operational concerns. The report includes a concise pull-out summary designed to serve the interests of operational forecast users.

The AMU's objective verification was originally designed to examine Meso-Eta model forecast errors over separate four-month periods from May through August 1996 (warm season) and from October 1996 through January 1997 (cool season). Given NCEP's ongoing development of the Eta model and the small sample sizes obtained from these limited four-month verification periods, the objective portion of the evaluation was extended to include secondary warm and cool season periods from May through August 1997 and October 1997 through January 1998, respectively. The twin-season comparison of forecast accuracy is helpful for model users by highlighting the model's characteristic strengths and weaknesses before and after the incorporation of model updates.

Results from the twin-season objective verification which can be important for operational forecast concerns include the following:

- The surface error characteristics vary widely by location, season, and time of day. The results are utilized most effectively by considering the model biases for each parameter separately and making appropriate adjustments to the forecasts.
- The random error component reveals substantial day-to-day variability in forecast accuracy. The random errors are caused primarily by the model's inability to resolve localized phenomena such as wind gusts, temperature gradients, or the effects of thunderstorms. While it is possible to partially adjust for model biases, it is much more difficult to accommodate the variability in forecast errors on any given day.
- Updates to the model's physical parameterizations produced identifiable and statistically significant changes in forecast accuracy at each location. Some changes enhanced forecast accuracy while others created larger systematic errors. It is important that model users maintain awareness of ongoing model changes. Such changes are likely to modify the basic error characteristics, particularly near the surface.
- On average, the forecast soundings at XMR and TBW during the warm season are too stable. The height of the lower tropospheric inversion at XMR and TBW was misrepresented during the cool season. Forecast biases for wind speed and direction are small at all three locations, but the random error component dominates the day-to-day variability. Given this variability, real-time assessment of forecast accuracy is necessary on any given day to help users determine if the model forecasts are consistent with current observations.
- The statistics did not reveal annual changes in upper-air forecast errors that could be attributed solely to February and August 1997 model updates. Moreover, since error growth is minimal, the error characteristics for upper-air forecasts apply, on average, at any time during the forecast period.
- The day-to-day fluctuations in convective indices are not well represented by the Meso-Eta model throughout the warm season. The convective index forecasts are most reliable overall during the cool season when, under normal circumstances, they provide little added value for most operational forecasting applications.

- Stratification of errors by the 950- to 600-mb layer-averaged wind direction does not reveal any substantial changes in error characteristics. Further, the stratification indicates that the day-to-day variations in the forecast errors may be more difficult to anticipate than general changes in the overall error characteristics under different wind regimes.

The AMU's statistical evaluation of Meso-Eta forecast accuracy identified a few biases that may result from inadequate parameterization of physical processes near the surface. Since the model bias or systematic error generally is small, most of the total model error results from day-to-day variability in the forecasts and/or observations. To some extent, these nonsystematic errors reflect the variability in point observations that sample spatial and temporal scales of atmospheric phenomena which cannot be resolved by the model. On average, Meso-Eta point forecasts may provide useful guidance for predicting the evolution of the larger scale environment. A more substantial challenge facing model users in real time is the discrimination of nonsystematic errors which could inflate the total forecast error.

While some of the changes in error characteristics due to model updates were expected, others were not consistent with the intent of the updates and further emphasize the need for ongoing sensitivity studies and localized statistical verification efforts. By pursuing ongoing localized verification efforts, model users could maintain an awareness of model updates and the effects that such changes have on point forecast accuracy within their area of responsibility.

1.0 Introduction

1.1 Motivation for Objective Point Forecast Verification

Weather support for ground and aerospace operations at the Kennedy Space Center (KSC) and Cape Canaveral Air Station (CCAS) requires accurate forecasts of winds, clouds, ceilings, fog, rain, lightning, and visibility. Numerical weather prediction models provide guidance for such forecasts by estimating the future values of these and other parameters. Specifically, weather observations are assimilated into the forecast model and integrated forward in time using dynamical equations of motion and other more empirical physical parameterizations. Throughout the integration, surface and upper-air weather variables are periodically extracted from the model at a point that corresponds geographically to the location of interest. These point forecasts are often used to help identify future changes in temperature, moisture, or winds that may contribute to the formation of adverse weather.

For several years, Model Output Statistics (MOS; Glahn and Lowry 1972; Carter et al. 1989) from numerical weather prediction models such as the National Center for Environmental Prediction (NCEP) Medium Range Forecast and Nested Grid Models have been used prevalently as sources of localized point forecast guidance. Given an adequately populated sample of runs in which the model configuration is not changed, MOS provides added value to the forecast process by statistically accounting for characteristic strengths and weaknesses in model forecasts at specific locations. However, NCEP is now entering an era where improvements in modeling capabilities are occurring so rapidly (McPherson 1994) that traditional MOS applications may no longer be appropriate for newer models. On the other hand, the combination of data assimilation techniques, refinements in model physics, and advances in computing efficiency (McPherson 1994) are enhancing the accuracy of deterministic model point forecasts.

In order to maximize the benefits of point forecast guidance within an environment of ongoing changes, it is helpful for both forecasters and model developers to maintain an objective awareness of a model's error characteristics at specific locations. For example, the development of local techniques to help correct identifiable model errors in real time could improve objective point forecast accuracy (e.g. Homleid 1995; Stensrud and Skindlov 1996; Baldwin and Hrebenach 1998). Moreover, periodic examination of model error characteristics could help developers diagnose and correct possible deficiencies in the model's physical parameterizations.

1.2 Precedent and Progress

In the spring of 1996, the Applied Meteorology Unit (AMU) began an evaluation of the NCEP 29-km Eta (Meso-Eta) model. The goal was to document the model's error characteristics for the U.S. Air Force 45th Weather Squadron (45WS), the National Weather Service (NWS) Melbourne (MLB), and the NWS Spaceflight Meteorology Group (SMG). The evaluation originally comprised objective and subjective verification strategies designed to measure overall forecast utility. The original evaluation was conducted over separate four-month periods from May through August 1996 (warm season) and from October 1996 through January 1997 (cool season).

Following the initial evaluation, Manobianco and Nutter (1997; hereafter MN97) concluded that the small sample sizes obtained from the four-month verification periods limited the quality of the objective verification results. Meanwhile, NCEP continues to update the configuration and physical parameterizations of the operational Eta model that could induce changes in forecast accuracy. For these reasons, the objective portion of the evaluation was extended for a second year to include additional warm and cool season periods from May through August 1997 and October 1997 through January 1998, respectively.

1.3 Applied Meteorology Unit Tasking

Under the Mesoscale Modeling Task (005), Subtask 2, the AMU evaluated the most effective ways to use the Meso-Eta model to meet 45WS, SMG and NWS MLB requirements (MN97). The evaluation methodology was determined by a technical working group consisting of several meteorologists and forecasters from the AMU, 45WS, SMG, and NWS MLB. Based on recommendations from the technical working group, the AMU

determined the data acquisition requirements, and designed and implemented the evaluation protocol. Using the existing resources and methodology established for Subtask 2, the objective portion of the Meso-Eta verification was extended for a second year under the Mesoscale Modeling Task (005), Subtask 6.

1.4 Report Objectives and Organization

The objectives of this extended statistical evaluation are to:

- Assess Meso-Eta point forecast accuracy at three specific locations important for SMG, 45WS, and NWS MLB operational concerns:
 - Edwards Air Force Base, California (EDW)
 - Shuttle Landing Facility, Florida (TTS)
 - Tampa International Airport, Florida (TPA)
- Determine the effect of ongoing model changes on local point forecast accuracy.

The TTS and EDW stations are selected because they are the primary and secondary landing sites for the Shuttle. The TPA site is chosen to compare model errors at two coastal stations on the eastern (TTS) and western (TPA) edge of the Florida peninsula. Forecast accuracy is evaluated statistically for the same forecast variables examined in the original evaluation (Table 1.1). Surface variables are examined as a function of time while upper-air variables are examined as a function of both time and height.

Table 1.1. Meso-Eta verification parameters.

Parameter	Levels
Mean sea-level pressure	--
Wind speed and direction	10 m
Temperature and dew point temperature	2 m
Wind speed and direction plus u , v components	Selected Levels
Wind speed and direction	1000-100 mb
Temperature	by
Mixing Ratio	25 mb increments
Geopotential Height	
Precipitable water (PWAT)	--
Convective available potential energy (CAPE)	--
Convective inhibition (CIN)	--
Lifted index (LIFT)	--
K index (KINX)	--
Ground relative helicity (HLCY)	0-3 km
Microburst day potential index (MDPI)	--
Thickness	1000-850 mb
Mean layer wind	850-500 mb
Mean layer relative humidity	850-500 mb

A brief overview of the Eta model and its configuration is presented in section 2. Procedures for data collection and statistical analysis are described in section 3. Complete results for surface, upper-air, and convective index forecasts are presented in sections 4, 5, and 6, respectively. Results for layer averaged quantities are discussed in section 7, while efforts to stratify statistical results by wind direction are discussed in section 8. A review of subjective model performance as documented by MN97 is presented in section 9. Finally, the report concludes in section 10 with a summary of results and recommendations for future work.

Readers not interested in studying the full details of all the statistical results may review the model performance summaries offered in sections 4.1, 5.1, and 6.1 and the conclusions in section 10. The performance summaries are designed to enhance the utility of Meso-Eta point forecast guidance in real-time forecast

operations. The summaries are therefore condensed for quick reference in Appendix E, in a mini booklet designed to be removed from this report, and on a diskette containing an HTML formatted summary for use on any desktop computer.

2.0 Eta Model Overview

The primary mesoscale modeling efforts at NCEP are focused on the development of the Eta model. A detailed summary of past Eta model development, including references, is provided in Appendix A. The following points are most important for operational users:

- The ongoing development of the Eta model will continue.
- NCEP implemented changes to the model's physical parameterizations midway through the AMU's 2-year objective evaluation. The statistics are stratified accordingly.
- The statistical evaluation was conducted on a version of the Eta model with 50 vertical levels and a 29-km horizontal grid spacing.
- The current operational version has 45 vertical levels and a 32-km horizontal grid spacing. This version is run four times per day.

3.0 Data and Analysis Method

3.1 Data Collection Periods

The AMU's objective and subjective verification was originally designed to consider 29-km Eta model forecast errors over separate four-month periods from May through August 1996 (warm season) and from October 1996 through January 1997 (cool season). Given the ongoing changes to the Eta model configuration and the small sample sizes obtained from these limited four-month verification periods, the objective portion of the evaluation was extended to include secondary warm and cool season periods from May through August 1997 and October 1997 through January 1998, respectively. The correspondence between these twin-seasonal evaluation periods and relevant Eta model updates is described in Table 3.1. The most substantial modifications were implemented in February 1997 at a time that falls between the 1996 and 1997 datasets. The timing of this update is convenient for the identification of changes in forecast accuracy, particularly for variables influenced by boundary layer processes.

Table 3.1. Definition of seasonal verification periods and notable Eta model updates.

Verification period	Date began	Date ended	Notable Eta model changes (EMC 1997)
1996 warm season	1 May 1996	31 August 1996	
1996 cool season	1 October 1996	31 January 1997	
1997 warm season	1 May 1997	31 August 1997	Radiation, cloud fraction, soil moisture, etc. (18 Feb. 1997)
1997 cool season	1 October 1997	31 January 1998	Corrected PBL depth computation (19 Aug. 1997)

3.2 Data Collection Methods

Forecasts from the 0300 UTC and 1500 UTC Meso-Eta model cycles were obtained via the internet from the National Oceanic and Atmospheric Administration's (NOAA) Information Center (NIC) ftp server. These files contain 33-h forecasts of surface and upper air parameters at 1-h intervals. NCEP extracts these surface and upper air station forecasts from the Meso-Eta model grid point nearest to the existing rawinsonde observation sites. Note that NCEP simply extracts the forecasts from the nearest grid point and does not

interpolate data from multiple surrounding points to the rawinsonde location. No attempt was made to correct for location or elevation errors that might exist in the forecast data. Instead, emphasis was placed on evaluating the raw operational product that is normally available in real time.

Hourly surface observations from TTS, TPA, and EDW are used to verify Meso-Eta surface forecasts. Upper-air forecasts are verified using available rawinsonde observations from EDW, Cape Canaveral Air Station (XMR), and Tampa Bay (TBW). Log-linear interpolation of data is used between reported pressure levels for verification at 25-mb intervals from 1000 to 100 mb. While surface forecasts are verified hourly, upper-air forecasts are verified only for those hours coinciding with the available rawinsonde release times. Surface and rawinsonde observation sites are not collocated at XMR and TBW, but the available sites are separated by not more than about 30 km (i.e. the Meso-Eta grid spacing). In order to avoid confusion, all subsequent references to surface and upper-air forecast verification will use only the rawinsonde station identifiers XMR, TBW, and EDW.

3.3 Statistical Analysis Methods

Statistical measures used to quantify Meso-Eta forecast errors (forecast – observed) include the

- Bias,
- Root Mean Square (RMS) error, and
- Error Standard Deviation.

By convention, all errors are defined by subtracting observations from forecasts. A positive bias indicates that the forecast variable is, on average, greater than observed. A negative bias indicates that the forecast variable is, on average, less than observed. The RMS error describes the overall magnitude of the total forecast error and is, by definition, a positive value. The error standard deviation is also a positive value and describes how widely the forecast errors fluctuate about the bias. This can be interpreted as the magnitude of day-to-day changes in forecast error. The statistical measures are defined mathematically in Appendix B.

For interpretation of results, it is helpful to recognize that the total model error (RMS) includes contributions from both systematic and nonsystematic sources. Systematic errors (model biases) are usually caused by a consistent misrepresentation of such factors as orography, radiation and convection. Nonsystematic errors are indicated by the error standard deviations and represent the random error component caused by initial condition uncertainty or inconsistent resolution of scales between the forecasts and observations.

In order to determine if model updates (Table 3.1) led to a statistically significant annual change in forecast accuracy, a Z statistic (Walpole and Meyers 1989) is calculated for a given parameter and compared with the normal distribution using a 99% confidence level. Additional details regarding statistical calculations are provided in Appendix B.

For quality control, gross errors in the data are screened manually and corrected, if possible. Errors that are greater than three standard deviations from the mean error (bias) are excluded from the final statistics. This procedure is effective at flagging bad data points and removes less than one percent of the data.

4.0 Surface Forecast Accuracy

Results from the comprehensive statistical evaluation of all variables listed in Table 1.1 are somewhat overwhelming upon initial review. Therefore, a summary of results is presented in section 4.1 with emphasis on operational interpretation. The summary is followed by a more complete presentation of results in sections 4.2 and 4.3, including a comparison of errors before and after changes in the model's physical parameterizations.

4.1 Overall Summary and Interpretation

Error characteristics for surface parameter forecasts vary widely by location, season, and time of day. The statistics can be utilized most effectively by considering the model biases for each parameter separately. For example, the fact that Meso-Eta wind speed forecasts are too fast on average at XMR (Table 4.1) suggests that

forecast accuracy might be improved by adjusting such guidance to lower speeds. Similar adjustments should be made to accommodate the biases identified for other parameters.

The random error component reveals substantial day-to-day variability in forecast accuracy. For many parameters, the random errors are larger than the corresponding biases, or systematic model errors. The random errors are caused primarily by the model's inability to resolve localized phenomena such as wind gusts, temperature gradients, or the effects of thunderstorms. While it is possible to partially adjust for model biases, it is much more difficult to accommodate the variability in forecast errors on any given day. It might help to compare current observations with the latest forecast guidance and make appropriate adjustments.

As numerical weather prediction systems such as the Meso-Eta model are updated with greater temporal and spatial resolution, they tend to exhibit smaller biases and larger error standard deviations. The nonsystematic, random error component associated with any model's inability to resolve local phenomena prevents perfect forecast guidance. However, the relatively minor biases indicate that on average, point forecasts provide useful guidance for the basic sensible weather variables considered here.

Results shown in sections 4.2 and 4.3 indicate that changes to the model's physical parameterizations produced identifiable and statistically significant changes in forecast accuracy at each location. Some changes enhanced forecast accuracy while others created larger errors. Since the model updates affected the basic error characteristics, statistics from the extended portion of the evaluation during 1997 are most representative of the model's current capabilities. Therefore, the error summaries presented in Tables 4.1–4.3 describe the Meso-Eta model's error characteristics using only the data collected during 1997. However, it is important that model users maintain awareness of ongoing model changes. Such changes are likely to modify the basic error characteristics, particularly near the surface.

Table 4.1. Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at XMR during the warm (May through Aug 1997) and cool (Oct 1997 through Jan 1998) seasons. A range of errors reveals fluctuations with time of day as demonstrated in sections 4.2 and 4.3.

Variable	Season	RMS	Bias	Std Dev	Interpretation
Sea-level Pressure (mb)	Warm	1	–1 to 0	1	Forecasts tend to be slightly lower than observed.
	Cool	1	±0.5	1	Small, variable forecast bias with random errors of 1 mb.
Temp. (°C)	Warm	1 to 2	±1	1 to 2	Forecasts are slightly warm in afternoon, slightly cool at night. Large random error component.
	Cool	2	0 to 1	2	Slight warm bias throughout the forecast cycle. Random error contributes more than bias.
Dew Point (°C)	Warm	1 to 2	–1 to 0	1 to 2	Forecasts are slightly dry on average. Random error contributes more than bias.
	Cool	1 to 3	0 to 2	1 to 2	Forecasts are typically wetter than observed.
Wind Speed (m s ^{–1})	Warm	2	0 to 2	1 to 2	Forecast winds are too fast on average.
	Cool	2 to 3	1 to 3	1.5	Forecast winds are too fast on average.
Wind Dir. (°)	Warm	50 to 70	±10	50 to 70	Forecasts are nearly unbiased although random errors are large.
	Cool	40 to 60	±10	40 to 60	Same as warm season except random errors are slightly smaller.

Table 4.2. Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at TBW during the warm (May through Aug 1997) and cool (Oct 1997 through Jan 1998) seasons. A range of errors reveals fluctuations with time of day as demonstrated in sections 4.2 and 4.3.

Variable	Season	RMS	Bias	Std Dev	Interpretation
Sea-level Pressure (mb)	Warm	1	–1 to 0	1	Forecasts tend to be slightly lower than observed.
	Cool	1	±0.5	1	Small, variable forecast bias with random errors of 1 mb.
Temp. (°C)	Warm	2.5	–3 to 1	1 to 2	Forecasts are too warm in the afternoon, too cool at night.
	Cool	1 to 3	–1 to 3	1 to 2	Forecasts are too warm in the afternoon, too cool at night.
Dew Point (°C)	Warm	1 to 2	–1 to 0	1 to 2	Forecasts are slightly dry on average. Random error contributes more than bias.
	Cool	1 to 3	0	1 to 3	Forecasts are unbiased but random errors reduce accuracy.
Wind Speed (m s ^{–1})	Warm	1.5	±1	1 to 2	Small forecast bias. Random error contributes more than bias.
	Cool	2	0 to 1	1.5	Forecast winds are slightly fast on average.
Wind Dir. (°)	Warm	50 to 80	–30 to 0	50 to 80	Forecast winds should be backed slightly to better match the observations.
	Cool	30 to 50	–20 to 0	30 to 50	Same as warm season except random errors are smaller.

Table 4.3. Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at EDW during the warm (May through Aug 1997) and cool (Oct 1997 through Jan 1998) seasons. A range of errors reveals fluctuations with time of day as demonstrated in sections 4.2 and 4.3.

Variable	Season	RMS	Bias	Std Dev	Interpretation
Sea-level Pressure (mb)	Warm	1 to 3	–2 to 0	1.5	Forecasts tend to be lower than observed.
	Cool	2 to 3	0 to 3	2	Forecasts tend to be greater than observed.
Temp. (°C)	Warm	3 to 6	–6 to –2	1 to 3	Forecasts are too cold on average.
	Cool	3 to 5	–4 to 0	2 to 4	Forecasts are too cold on average, especially during the daytime.
Dew Point (°C)	Warm	3 to 9	0 to 8	3 to 5	Forecasts are too moist on average, especially during the daytime.
	Cool	3 to 6	–1 to 5	3.5	Forecasts are mostly wetter than observed, especially during the daytime.
Wind Speed (m s ^{–1})	Warm	2 to 6	–7 to –1	1.5 to 3	Forecasts too slow on average, especially during the daytime.
	Cool	2 to 3	–2 to 0	2	Forecasts too slow on average.
Wind Dir. (°)	Warm	20 to 90	0 to 30	20 to 90	Forecast winds should be veered slightly overnight to better match the observations.
	Cool	60 to 90	0 to 30	60 to 90	Same as warm season.

4.2 1996 Surface Forecast Accuracy

In the following section, Meso-Eta point forecast error characteristics for the surface variables in Table 1.1 are examined in greater detail for both the 1996 warm and cool seasons. Although statistics were examined separately for the 0300 and 1500 UTC forecast cycles, only those from the 0300 UTC cycle are shown here. Results from the 1500 UTC cycle provide little additional information since positive or negative biases occur with comparable magnitudes at approximately the same time of day in both forecast cycles. Combining data from both the 0300 and 1500 UTC cycles as a function of forecast duration tends to cancel out the diurnally varying errors. For operational concerns, the error characteristics described here for the 0300 UTC forecast cycle apply equivalently for the 1500 UTC cycle.

4.2.1 1996 Warm Season

4.2.1.1 Mean Sea-Level Pressure

During the 1996 warm season, biases in mean sea level pressure are less than ± 1 mb at XMR and TBW (Fig 4.1a). At EDW, the bias varies throughout the forecast period and reaches a maximum of nearly 4 mb around 1300 UTC. Since RMS errors and error standard deviations are comparable in magnitude at XMR and TBW (Figs. 4.1b, c), much of the total error for these locations evidently is nonsystematic in nature. Conversely, the large biases at EDW contribute strongly to the RMS error and therefore represent a systematic model error. One possible explanation for this apparent model deficiency at EDW may be that the forecast point data extracted from the model are almost 250 m lower than the actual station elevation.

4.2.1.2 2-m Temperature

Warm season biases in 2-m temperature at XMR and TBW follow a diurnal cycle as values range from about -3 to 1 °C (Fig. 4.1d). The amplitude of the diurnal cycle is larger at EDW, with cold biases reaching almost -6 °C during the early part of the forecast. Since forecast biases and corresponding RMS errors are comparable in magnitude at EDW (Figs. 4.1d, e), the larger contribution to the total error for this location evidently is derived from a systematic model error. This model error at EDW could be related to the incorrect specification of station elevation. The results at all three locations are also consistent with those from Betts et al. 1997 and Black et al. 1997 (hereafter BE97 and BL97) who found an excessive range of summer temperatures due to radiation errors in the 1996 version of the 48-km Eta model.

4.2.1.3 2-m Dew Point Temperature

Warm season biases in 2-m dew point temperature at XMR and TBW are generally smaller than ± 2 °C (Fig. 4.1g). Biases at EDW are positive (moist) during the first 21 h of the forecast cycle (Fig. 4.1g). When viewed in conjunction with the 2-m temperature bias in Fig. 1a, the net result is that forecasts at EDW are too cold and moist over this period.

The studies by BE97 (their Fig. 10b) and BL97 (their Fig. 4b) indicate excessive amounts of 2-m specific humidity in the forecasts at time zero using regionally averaged data during the summer. Their results also reveal that after time zero, specific humidity levels are underforecast on average throughout the remainder of the forecast cycle. Here, zero-hour dew point errors at EDW are consistent with results from those studies but the enduring positive bias indicates clearly that regionally averaged statistics can mask important error characteristics that are specific to particular locations. Some of the difficulties in forecasting dew point temperatures at EDW could relate to problems with PBL mixing and/or incorrect specification of soil moisture processes as discussed by BE97. Such difficulties would likely be exacerbated by the station elevation error at EDW and by post-processor errors while translating mixing ratios into 2-m dew point temperatures.

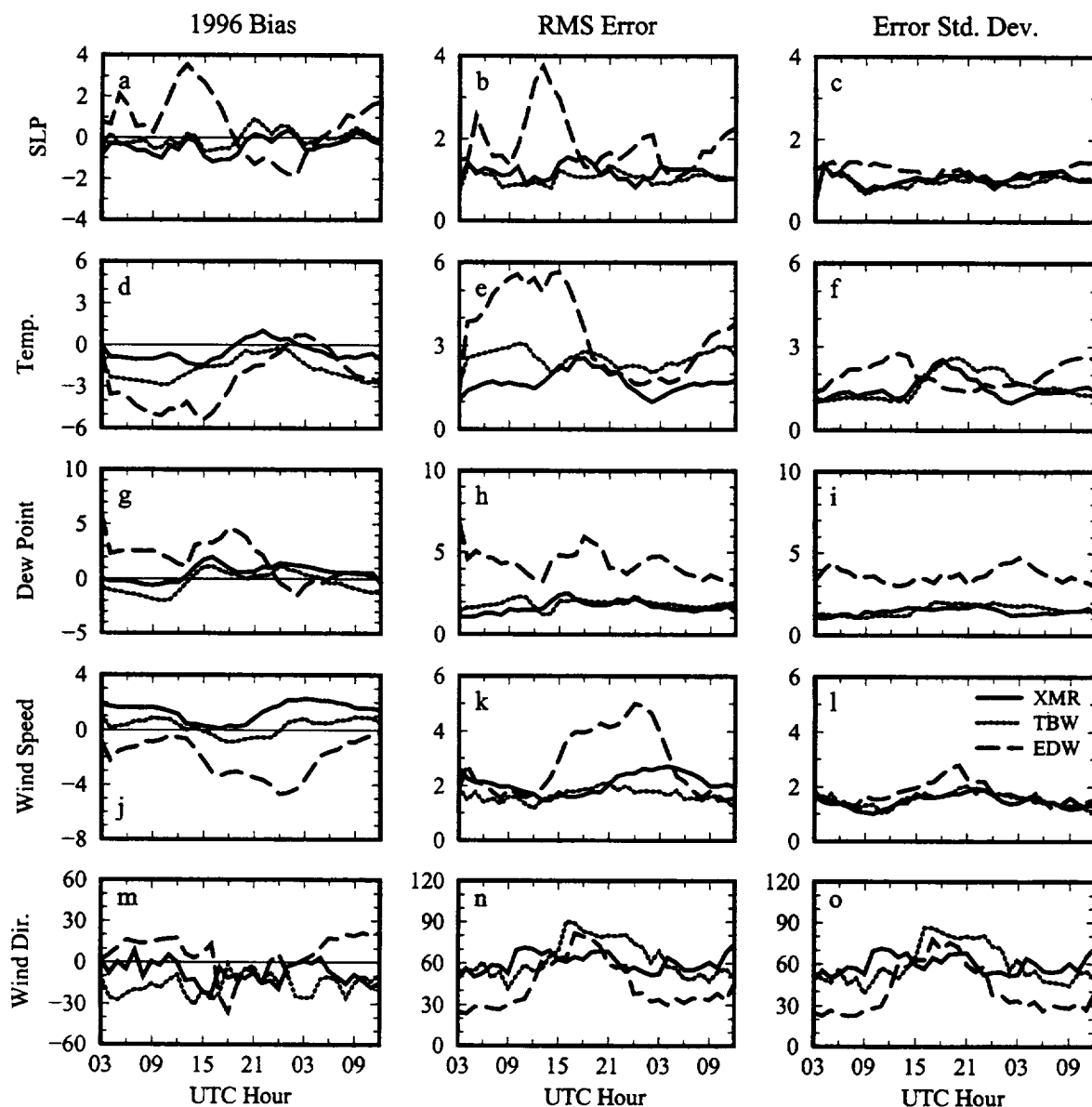


Figure 4.1. Bias (forecast – observed), RMS error, and error standard deviation for surface parameter forecasts from the 0300 UTC Meso-Eta cycle during the 1996 warm season. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Statistics for mean sea-level pressure (mb), 2-m temperature and dew point temperature ($^{\circ}\text{C}$), and 10-m wind speed (m s^{-1}) and direction ($^{\circ}$) are shown respectively in panels a-c, d-f, g-i, j-l and m-o.

4.2.1.4 10-m Wind Speed

Warm season biases in 10-m wind speed range from 0 to -5 m s^{-1} at EDW and from -1 to 2 m s^{-1} at XMR and TBW (Fig. 4.1j). Therefore, 10-m wind speed forecasts at XMR and TBW tend to be slightly fast on average while those at EDW are generally too slow. The relatively large increase in the magnitudes of biases and RMS errors at EDW between about 1500 and 0300 UTC reflects a period during which systematic model errors comprise the larger portion of the total forecast error (compare Figs. 4.1j-l).

4.2.1.5 10-m Wind Direction

Warm season biases in 10-m wind direction vary within about $\pm 30^\circ$ at XMR and EDW (Fig 4.1m). The forecast wind direction at TBW is, on average, counterclockwise (negative) relative to the observed wind direction. The RMS errors range from about 30 to 90° , with the largest errors occurring during the middle of the forecast period (Fig 4.1n). Since the biases are small compared to the error standard deviation, much of the wind direction error is determined by nonsystematic sources. Since the model cannot temporally or spatially resolve many local effects which influence wind direction such as topography or vegetation, the magnitude of variability in wind direction errors is not surprising especially when wind speeds are light.

4.2.2 1996 Cool Season

4.2.2.1 Mean Sea-Level Pressure

During the 1996 cool season, biases in mean sea-level pressure fluctuate from about 1 to -2 mb at XMR and TBW (Fig 4.2a). The bias is largest at EDW, where mean errors remain steady around 2 mb throughout much of the forecast period. Since the RMS errors and error standard deviations at XMR and TBW are comparable in magnitude during most of the forecast cycle, nonsystematic errors evidently provide a strong contribution to the total error (Figs. 4.2a, b). At EDW however, it is not clear whether systematic or nonsystematic errors contribute more to the total error for that location (compare Figs. 4.2a-c).

4.2.2.2 2-m Temperature

During the 1996 cool season, 2-m temperature biases are slightly positive at XMR and slightly negative at TBW, with errors ranging from about 0 to 2°C and 0 to -2°C , respectively (Fig. 4.2d). Forecast temperatures at EDW are about 0 to -4°C colder than observed on average. Over the first 12 h of the forecast cycle, large error standard deviations at EDW (Fig. 4.2f) suggest that nonsystematic errors contribute to a substantial portion of the total model error. During the middle part of the forecast cycle from about 1500 to 0300 UTC, the larger negative bias at EDW indicates that systematic model errors contribute more strongly to the total error. Cool season temperature errors at TBW have nearly the same characteristics as those from the previous warm season. At XMR and EDW however, the cool season results do not clearly show diurnal fluctuations that would otherwise be consistent with an excessive range of temperatures in the 1996 Eta model configuration. Additional sensitivity studies are therefore necessary in order to determine other possible sources of systematic model error that might degrade the accuracy of temperature forecasts during the cool season.

4.2.2.3 2-m Dew Point Temperature

Cool season biases in 2-m dew point temperature at all three stations are generally larger than those of the previous warm season (compare Figs. 4.1g, 4.2g). Biases at TBW range from about -1 to 3°C while at XMR, a moist bias of 3 to 4°C is evident throughout much of the forecast cycle. Qualitatively, the difference in error characteristics at XMR and TBW is notable given their relative proximity. Model biases at EDW follow similar fluctuations with time during both seasons, but reach slightly higher maximum values of around 6°C during the cool season at 2100 UTC. Difficulties remain at EDW during the cool season for initializing the zero-hour dew point temperatures. The overall cool season increase in forecast biases contributes to a corresponding growth in RMS error at all three locations (Fig. 4.2h). This result suggests that systematic errors in Eta model dew point temperature forecasts are larger during the cool season.

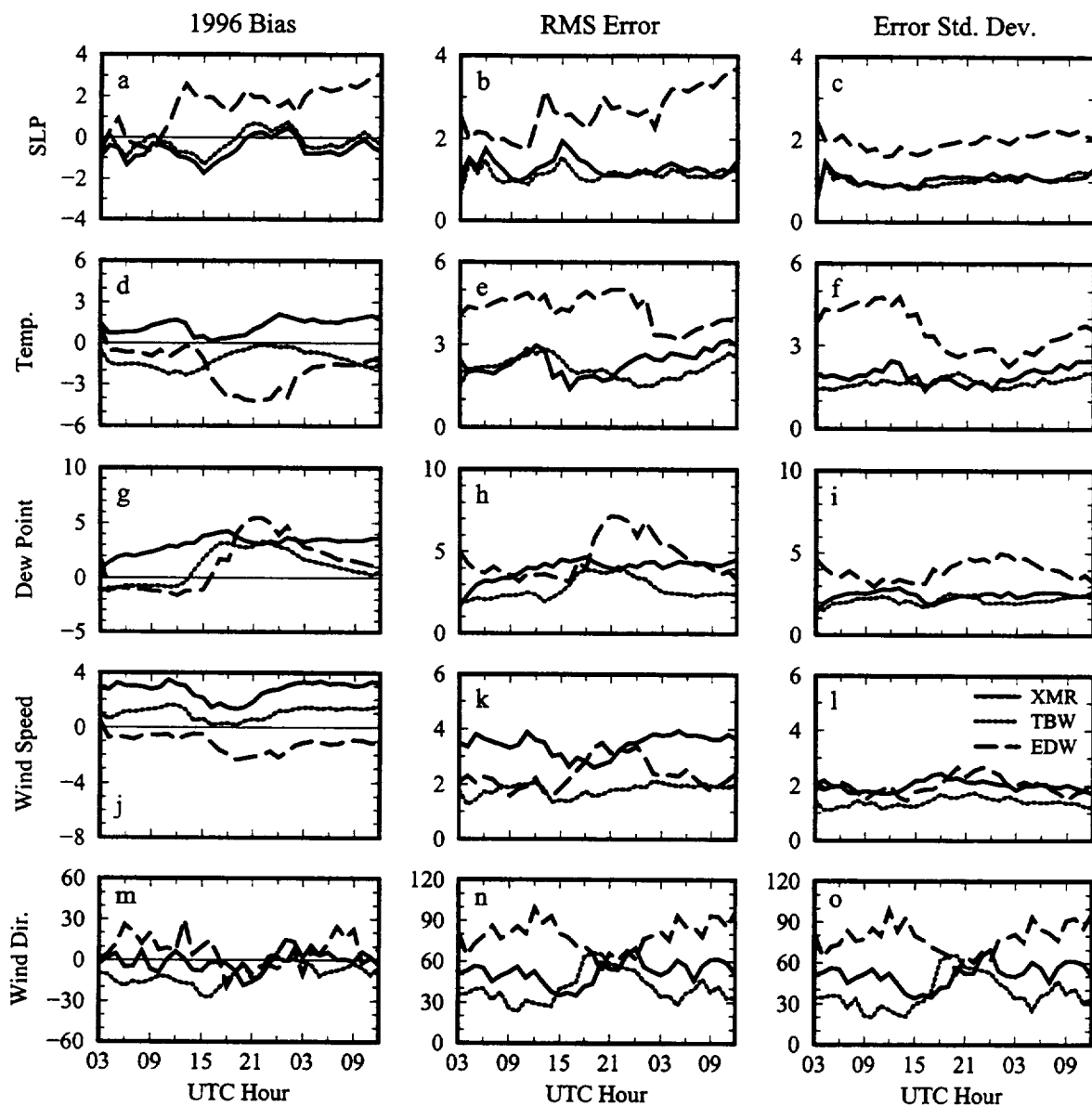


Figure 4.2. Bias (forecast – observed), RMS error, and error standard deviation for surface parameter forecasts from the 0300 UTC Meso-Eta cycle during the 1996 cool season. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Statistics for mean sea-level pressure (mb), 2-m temperature and dew point temperature ($^{\circ}\text{C}$), and 10-m wind speed (m s^{-1}) and direction ($^{\circ}$) are shown respectively in panels a-c, d-f, g-i, j-l and m-o.

4.2.2.4 10-m Wind Speed

Cool season wind speed biases at XMR are about 1 m s^{-1} greater than those during the warm season (compare Figs. 4.2j, 4.1j). The combined cool season increases in forecast biases and error standard deviations at XMR result in RMS errors which are about 1.5 m s^{-1} larger than corresponding warm season errors (Figs. 4.2k, l). Wind speed biases at TBW are comparable during both seasons while the slow bias at EDW improves in the cool season.

4.2.2.5 10-m Wind Direction

Cool season errors in 10-m wind direction are similar to those of the previous warm season. Biases again are less than $\pm 30^\circ$ at all three locations (Fig. 4.2m). The RMS errors and error standard deviations at XMR and TBW range from 30 to 70° (Figs. 4.2n, o). At EDW, these errors are slightly larger relative to their warm season values. The greater portion of the total error at all three stations remains nonsystematic in nature (Fig 4.2o).

4.3 Changes in Surface Forecast Accuracy During 1997

The AMU's original Meso-Eta evaluation (MN97) was extended, in part, to enhance the quality of results by increasing sample sizes. However, as described below, the model upgrades implemented between the 1996 and 1997 evaluation periods (Table 3.1) produced many statistically significant changes in forecast accuracy. For this reason, the data are not combined into a single large sample for evaluation. Instead, the analysis is repeated here for the 1997 data with emphasis on evaluating changes in forecast accuracy between 1996 and 1997. A comparison of results between the two seasons highlights the necessity for model users to maintain an awareness of forecast accuracy at specific locations in lieu of ongoing changes.

The following points are considered while evaluating changes in surface forecast accuracy during 1997 at XMR, TBW, and EDW.

- Model biases during 1997 are presented to provide an updated assessment of mean forecast accuracy following the model changes. These results were summarized earlier in Tables 4.1–4.3.
- Annual changes in the absolute value of model biases ($|1997 \text{ bias}| - |1996 \text{ bias}|$) are examined to determine whether the model systematic errors became larger or smaller relative to zero. The use of absolute value assumes, for example, that a cold bias of 2 degrees is as serious for operational concerns as a warm bias of 2 degrees. More generally, a positive difference reveals that the systematic error is larger, or farther from zero, and that the forecasts are on average more biased in 1997. A negative difference reveals that the systematic error is smaller, or closer to zero, and that the forecasts are on average less biased in 1997.
- Whenever changes in forecast biases are statistically significant (Appendix B), efforts are made to determine whether the changes may be attributed to annual differences in either the mean forecasts or observations. Whenever bias changes are explained largely by differences in mean forecast values, it is likely that the model updates led to an improvement or degradation in forecast accuracy during 1997. Otherwise, the difference simply reflects the interannual variability in the observations.

4.3.1 1997 Warm Season

4.3.1.1 Mean Sea-Level Pressure

During the 1997 warm season, biases in mean sea-level pressure range from about -1 to 0 mb at XMR and TBW while at EDW, the pressure is underforecast by about -3 to 0 mb (Fig. 4.3a). The annual change in pressure biases at XMR and TBW was nearly zero (Fig. 4.3b). The systematic error at EDW improved by 3 mb around 1300 UTC while, at other times of day, the error increased by about 2 mb. The standardized Z statistic indicates that the annual changes in bias at EDW are statistically significant at the 99% confidence level (Fig. 4.3c). The observed sea-level pressure at XMR and TBW during the 1997 warm season was, on average, about 1.5 mb lower than the corresponding 1996 average (Fig. 4.4b). Since the model responded correctly to this average decrease in pressure (Fig. 4.4a), the annual change in forecast error was insignificant at XMR and TBW (Fig. 4.3c). At EDW however, the model failed to capture the increase in the average sea-level pressure observed for that location in 1997 (Figs 4.4a, b). Since the average forecast pressure remained nearly unchanged at EDW during 1996 and 1997, it is not clear that the February 1997 model updates contributed to the changes in pressure forecast errors.

4.3.1.2 2-m Temperature

The 2-m temperature biases at XMR and TBW range from about -3 to 1 °C while at EDW, forecasts are on average 2 to 6 °C colder than observed throughout much of the forecast cycle (Fig. 4.3d). The annual change in temperature biases at XMR and TBW was nearly zero (Fig. 4.3e). At EDW, the systematic error increased by about 3 °C between 1500 and 0600 UTC. The standardized Z statistic indicates that these larger biases at EDW are statistically significant at the 99% confidence level (Fig. 4.3f). Average forecast temperatures at XMR and TBW increase slightly during 1997 while those at EDW are reduced by about -3 °C (Fig. 4.4c). Observed temperature climatologies are nearly identical at all three locations during both 1996 and 1997 (Fig. 4.4d). These results confirm that the stronger cold bias at EDW during the 1997 warm season is driven mostly by a reduction in forecast temperatures. The reduction of forecast temperatures at EDW during 1997 exacerbates an existing cold bias and contributes to a loss of accuracy for that location. These lower temperatures are consistent with the systematic decrease in the model's incoming solar radiation imposed in February 1997.

4.3.1.3 2-m Dew Point Temperature

The 2-m dew point temperature biases at XMR and TBW are slightly underforecast by about -1 °C during the 1997 warm season (Fig. 4.3g). Results at EDW continue to indicate a large positive (moist) bias in the forecasts at time zero and from about 1500 to 0300 UTC. Annual changes in the errors at XMR and TBW are less than ± 1 °C (Fig. 4.3h). At EDW, the change in absolute bias reveals enhanced accuracy over the first part of the forecast cycle, followed by an increase in error that reaches nearly 6 °C. The Z statistic (Fig. 4.3i) confirms that the annual changes in 2-m dew point temperatures are statistically significant during the middle of the forecast cycle at all three stations. The results shown in Fig. 4.4e indicate that these annual changes in bias are driven mostly by an increase (decrease) in the mean forecast values at EDW (XMR and TBW). By comparison, relatively minor shifts are noted in the average dew point temperature observations (Fig. 4.4f).

The Eta model updates implemented in February 1997 were designed to reduce PBL mixing and thereby improve the summer dry bias noted in specific humidity forecasts (BE97; BL97). Although increased values for 2-m dew point temperature forecasts at EDW (Fig. 4.4e) are consistent with the intent of these model updates, the change exacerbates an existing moist bias. The decreased moisture in the forecasts at XMR and TBW during 1997 is not expected and cannot be explained from this limited evaluation. However, since the annual change in systematic error for these locations is less than ± 1 °C (Fig. 4.3h), forecast utility should not be affected.

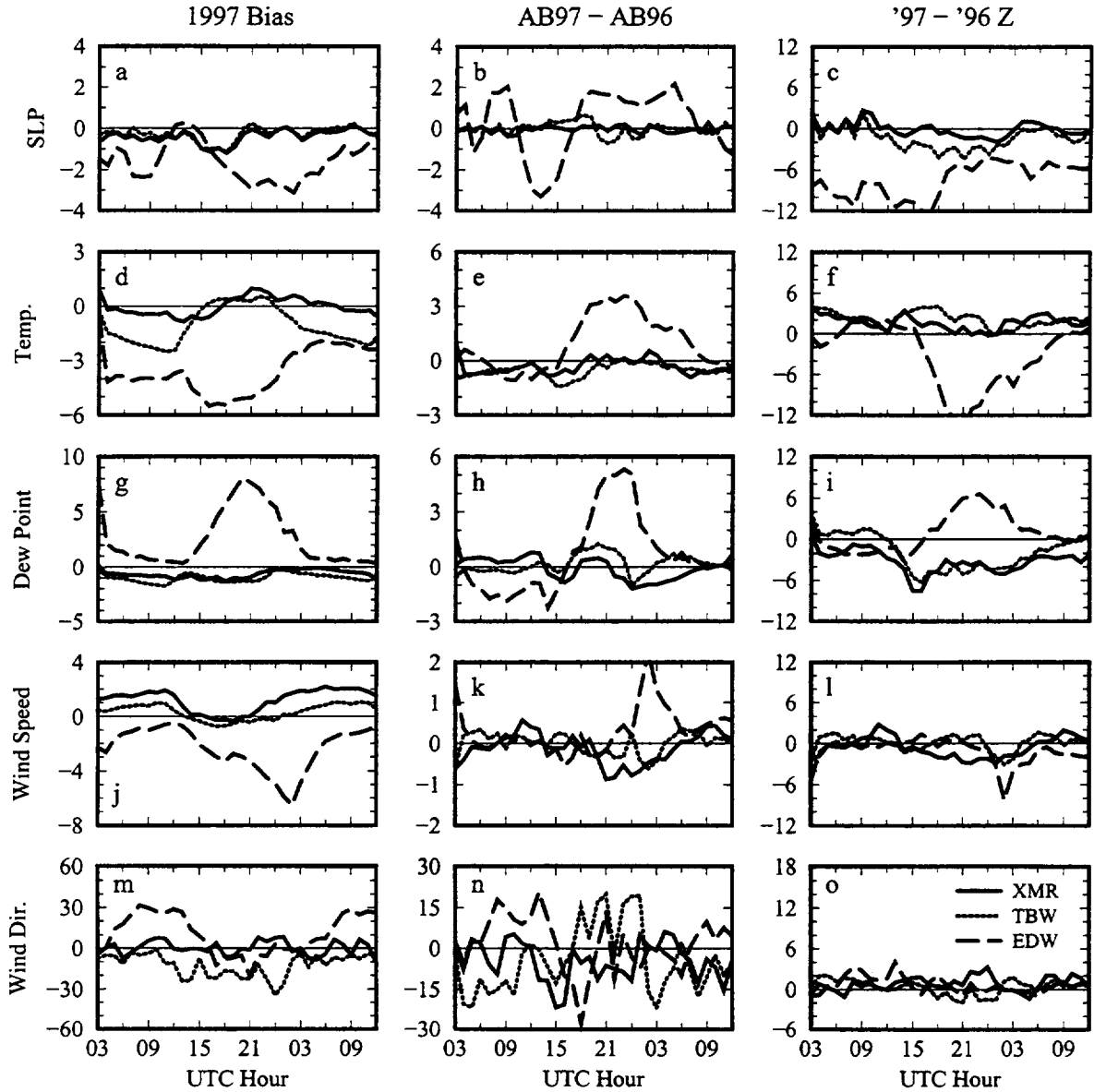


Figure 4.3. 1997 warm season bias (forecast - observed), annual difference of absolute bias (AB97 - AB96), and standardized Z statistics for the 0300 UTC Meso-Eta cycle. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Statistics for mean sea-level pressure, 2-m temperature and dew point temperature, and 10-m wind speed and direction are shown respectively in panels a-c, d-f, g-i, j-l and m-o. Units are mks except for the nondimensional Z statistic. Z scores that lie outside the shaded region indicate that changes between 1997 and 1996 warm season forecast biases are statistically significant at the 99% confidence level (see Appendix B).

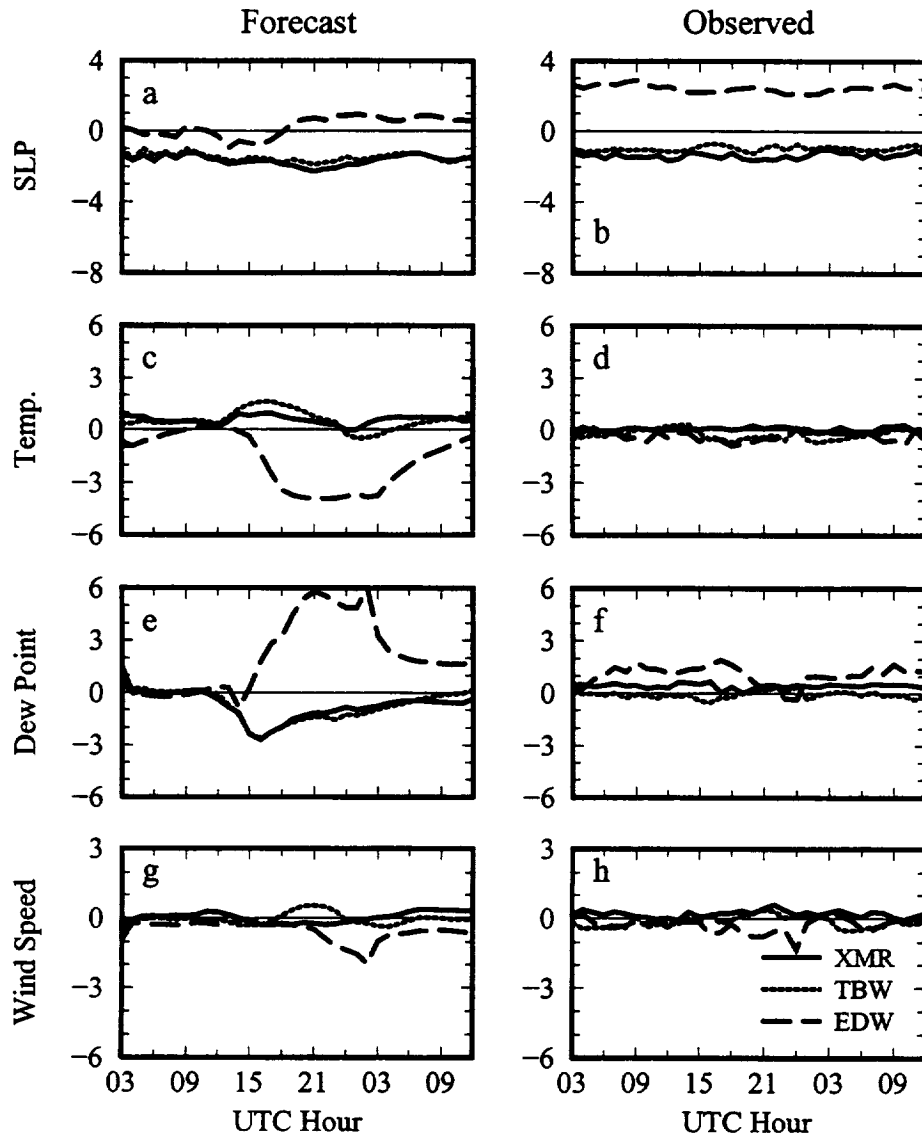


Figure 4.4. Comparison of annual changes (1997 - 1996) in mean forecasts and observations during the warm season. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Annual differences for average sea-level pressure (mb), 2-m temperature and dew point temperature ($^{\circ}\text{C}$), and 10-m wind speed (m s^{-1}) are shown respectively in panels a-b, c-d, e-f, and g-h.

4.3.1.4 10-m Wind Speed

Wind speed forecasts during the 1997 warm season are slightly fast at XMR and TBW with a continued slow bias at EDW (Fig. 4.3j). The only statistically significant annual changes occur at EDW around 0200 UTC where the magnitude of the biases increase by 2 m s^{-1} during 1997 (Figs. 4.3k, l). It is not clear whether these changes in bias are driven by model updates alone since the differences between 1996 and 1997 mean forecasts and observations are both small (Figs. 4.4g, h). This result is not surprising since the Eta model updates implemented in February 1997 were not designed explicitly to alter the forecast wind fields.

4.3.1.5 10-m Wind Direction

Wind direction biases fluctuate within $\pm 30^\circ$ during the 1997 warm season (Fig. 4.3m). The annual change in systematic error between 1996 and 1997 was less than 30° at all three locations (Fig. 4.3n). The standardized Z statistic reveals that these annual changes in mean error are not statistically significant at the 99% confidence level. Again, this result is not surprising since the Eta model updates implemented in February 1997 were not designed explicitly to alter the forecast wind fields.

4.3.1.6 Summary of 1997 Warm Season Changes

The Eta model updates implemented during February 1997 were designed to decrease low-level temperatures and increase the low-level moisture. The results shown above demonstrate the following changes in forecast biases at XMR, TBW, and EDW.

- Sea-level pressure, wind speed and wind direction biases did not change in response to internal model changes. Note that the model updates were not designed to affect these parameters.
- The existing cold, moist bias in temperature and dew point temperature forecasts at EDW became worse in 1997.
- Temperature and dew point temperature biases at XMR and TBW were relatively unaffected by the model changes.

4.3.2 1997 Cool Season

4.3.2.1 Mean Sea-Level Pressure

During the 1997 cool season, biases in sea-level pressure forecasts at XMR and TBW are less than ± 1 mb (Fig. 4.5a). At EDW, the forecast pressures are, on average, 1 to 2 mb greater than observed. The systematic error at all three locations decreased by about 0 to 1 mb during the 1997 cool season (Fig. 4.5b). Although the annual changes in error are not statistically significant at EDW, the Z score does reveal a few hours where the changes are significant at XMR and TBW (Fig. 4.5c). Examination of the annual differences in mean forecast and observed pressures (Figs. 4.6a, b) does not indicate clearly whether the significant changes in bias at XMR and TBW could be attributed to a change in either the forecasts or observations.

4.3.2.2 2-m Temperature

The 1997 cool season 2-m temperature forecasts at XMR are on average about 1°C warmer than observed (Fig. 4.5d). At EDW, forecasts are again colder than observed throughout much of the forecast cycle, especially from about 1500 to 0300 UTC. Biases at TBW indicate that the diurnal range of 2-m temperatures is overforecast slightly with values ranging from about -1 to 2°C .

The systematic errors at XMR and EDW are comparable during both 1996 and 1997 cool seasons with annual changes of less than $\pm 1^{\circ}\text{C}$ (Fig. 4.5e). At TBW however, the magnitude of the bias in 1997 increases by about 2°C during the middle of the forecast period. The statistical significance of this change is supported by large Z scores (Fig. 4.5f).

The increases in mean forecast temperatures at TBW during 1997 are larger than changes in the mean observed temperatures (Figs. 4.6c, d). But an increase in mean forecast temperatures during local daytime hours is not consistent with the intent of the February or August 1997 model updates and actually leads to a degradation of forecast accuracy at TBW. Notably, temperature biases at EDW are nearly identical during both cool seasons whereas annual differences in warm season data suggest a strong response to the February 1997 model updates. The statistics shown here, in combination with knowledge of changes to the model's physical parameterizations, evidently are not adequate to fully explain the source of all changes in systematic model errors.

4.3.2.3 2-m Dew Point Temperature

The 2-m dew point temperature bias during the 1997 cool season is less than 2°C at XMR and less than $\pm 1^{\circ}\text{C}$ at TBW (Fig. 4.5g). However, results at EDW continue to indicate a large positive (moist) bias in the forecasts at time zero and during the latter portions of the forecast cycle. While dew point temperature biases at EDW are similar during both 1996 and 1997 cool seasons, the systematic errors at XMR and TBW decrease by about 3°C (Fig. 4.5h). The Z statistic reveals that these annual changes in bias at XMR and TBW are statistically significant at the 99% confidence level (Fig. 4.5i). The enhanced forecast accuracy at XMR and TBW evidently results from a combination of lower (drier) dew point temperature forecasts and higher (wetter) observations on average during 1997 (Figs. 4.6e, f). In spite of the forecast improvements, these findings are not anticipated since the February and August 1997 model updates were designed to raise near-surface moisture levels (BE97; BL97).

4.3.2.4 10-m Wind Speed

Wind speed forecasts during the 1997 cool season remain too fast at XMR and TBW and too slow at EDW (Fig. 4.5j). The greatest annual change in systematic error occurs at XMR where the bias is reduced by at most 1 m s^{-1} (Fig. 4.5k). The Z scores shown in Fig. 4.5l confirm that no statistically significant changes occur in wind speed biases between the 1996 and 1997 cool seasons. Moreover, differences in mean forecast and observed wind speed between 1996 and 1997 are similar (Figs. 4.6g, h). As during the warm season, this cool season result is expected since the Eta model updates were not explicitly designed to modify wind speed forecasts.

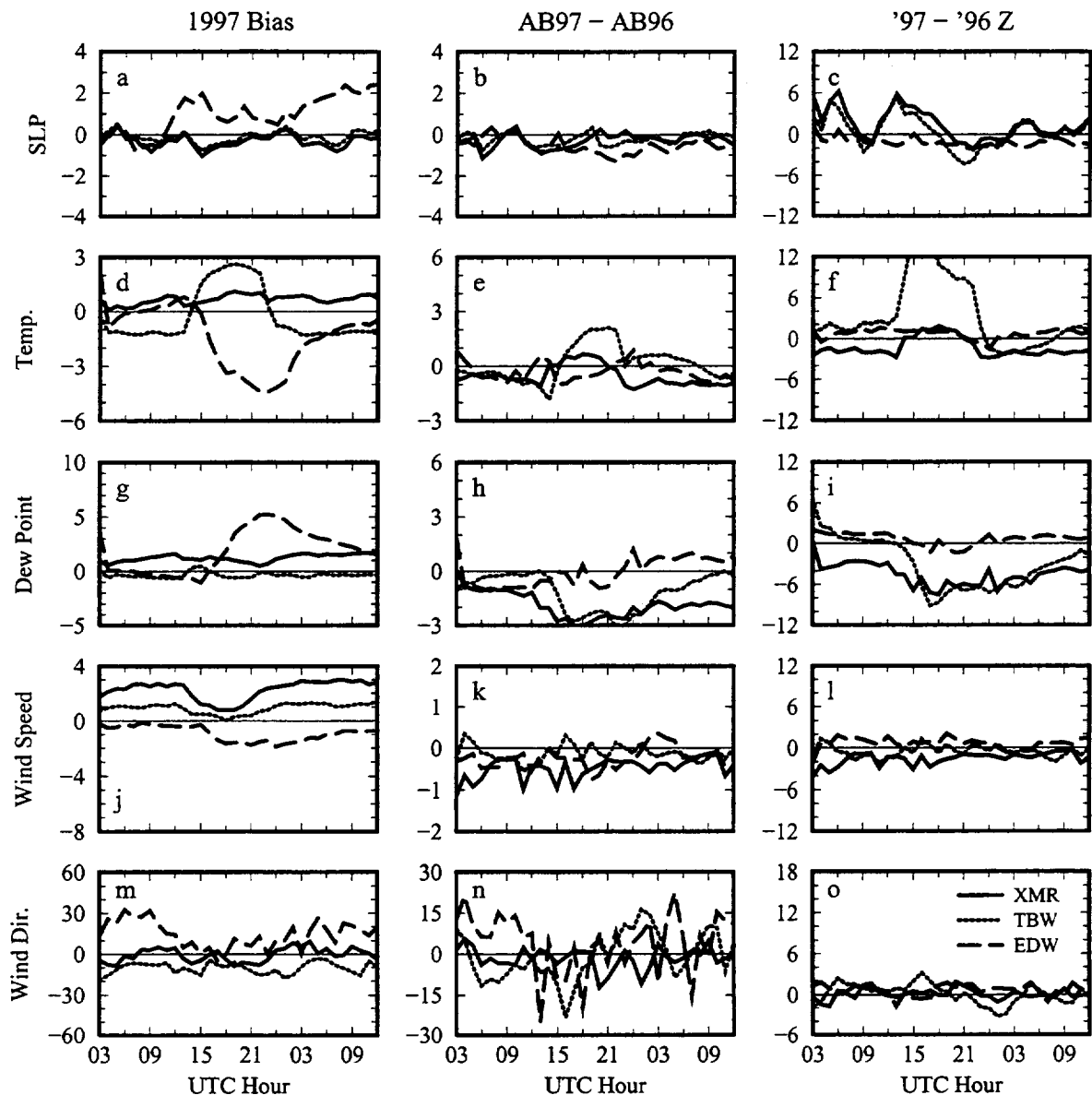


Figure 4.5. 1997 cool season bias (forecast – observed), annual difference of absolute bias (AB97 – AB96), and standardized Z statistics for the 0300 UTC Meso-Eta cycle. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Statistics for mean sea-level pressure, 2-m temperature and dew point temperature, and 10-m wind speed and direction are shown respectively in panels a-c, d-f, g-i, j-l and m-o. Units are mks except for the nondimensional Z statistic. Z scores that lie outside the shaded region indicate that changes between 1997 and 1996 warm season forecast biases are statistically significant at the 99% confidence level (see Appendix A).

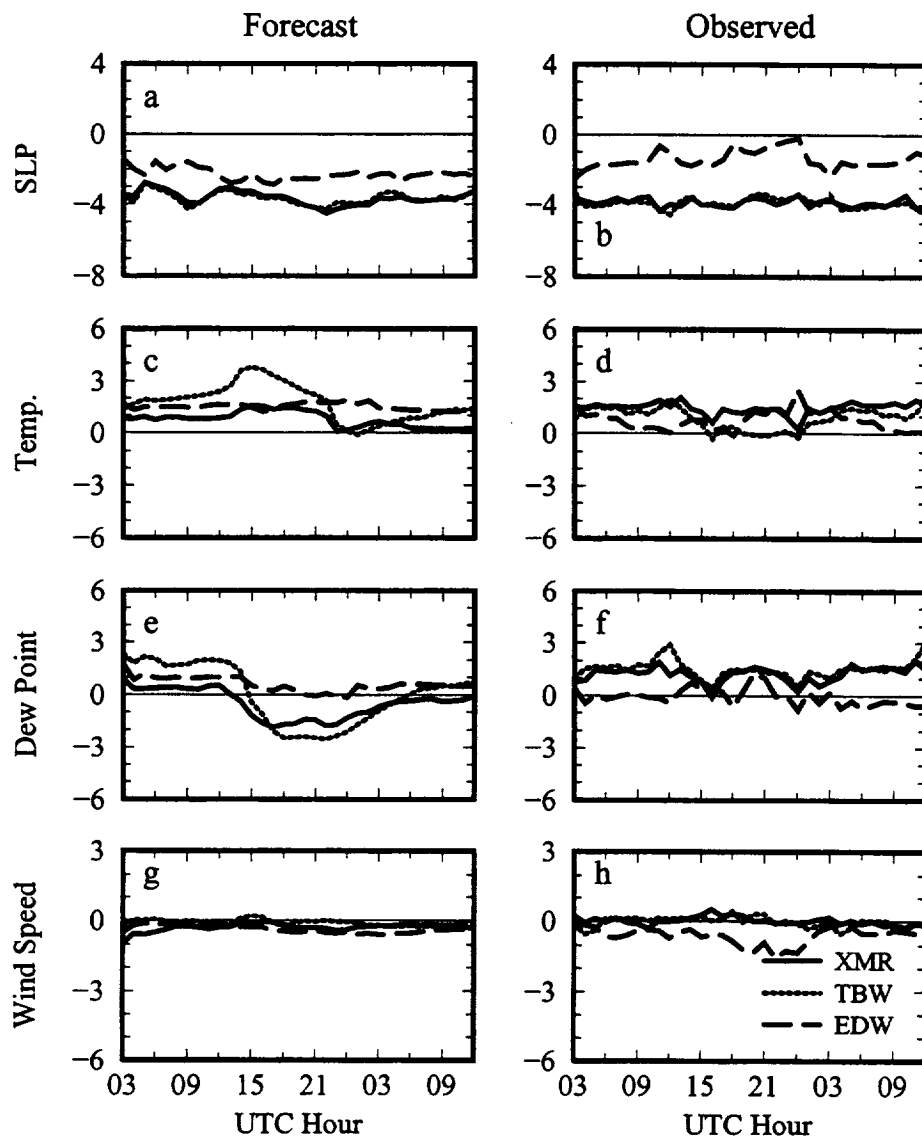


Figure 4.6. Comparison of annual changes (1997 - 1996) in mean forecasts and observations during the cool season. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Annual differences for average sea-level pressure (mb), 2-m temperature and dew point temperature ($^{\circ}\text{C}$), and 10-m wind speed (m s^{-1}) are shown respectively in panels a-b, c-d, e-f, and g-h.

4.3.2.5 10-m Wind Direction

The annual change in cool season wind direction bias is less than $\pm 20^\circ$ at all three locations (Fig. 4.5n). The standardized Z statistic reveals that none of the changes are statistically significant at the 99% confidence level (Fig. 4.5o). Again, these results are not surprising since the Eta model updates were not designed to explicitly modify the wind fields.

4.3.2.6 Summary of 1997 Cool Season Changes

The Eta model updates implemented during February and August 1997 were designed to decrease low-level temperatures and increase the low-level moisture. The results shown above demonstrate the following changes in forecast biases at XMR, TBW, and EDW.

- Sea-level pressure, wind speed and wind direction biases did not change in response to internal model changes. Note that the model updates were not designed to affect these parameters.
- A daytime warm bias was introduced for temperature forecasts at TBW. This increase in error was not anticipated since model changes were designed to reduce temperatures.
- There was no change in the error characteristics for temperature forecasts at EDW during the cool season. This is in contrast to the large reduction in temperature noted during the previous warm season.
- The systematic error in dew point temperature forecasts was reduced at XMR and TBW. However, the statistics do not clearly indicate the source of such change.

4.4 10-m Wind Persistence

The only accuracy benchmark specified in the original evaluation protocol was a comparison of 10-m winds with 1 to 6 h persistence. Since the February and August 1997 model updates did not produce a significant impact on the accuracy of wind forecasts, the reader is referred to MN97 for this comparison of wind forecasts with persistence. In general, they found that 1- to 3-h persistence forecasts of wind speed and direction usually have smaller RMS errors than the corresponding Meso-Eta model forecasts. However, the model forecasts of these variables are occasionally more accurate than 6-h persistence.

5.0 Upper Air Forecast Accuracy

The AMU's original Meso-Eta evaluation (MN97) was extended, in part, to enhance the quality of results by increasing sample sizes. For the surface parameters discussed previously in section 4, it was not reasonable to combine data from 1996 and 1997 because model updates produced identifiable and statistically significant changes in forecast accuracy. However, examination of error statistics for the upper-air forecasts at XMR, TBW, and EDW reveals only subtle changes in their characteristics between 1996 and 1997 (annually stratified results not shown). This is not surprising since the Eta model changes implemented in February and August 1997 (Table 3.1) were designed primarily to improve deficiencies in forecasts for surface and boundary layer variables (BE97, EMC 1997). For these reasons, all upper-air data collected during 1996 and 1997 are pooled into their respective warm and cool season periods to develop generalized profiles of Meso-Eta error characteristics at XMR, TBW, and EDW.

A summary of results is presented in section 5.1 with emphasis on operational interpretation. The summary is followed by a more complete description of upper-air forecast accuracy in section 5.2 and an evaluation of forecast error growth in section 5.3.

In addition to the graphical results in section 5.2, tables are presented in Appendix C which list errors in u and v wind components, wind speed, and wind direction as a function of height and verification period for XMR and EDW. These statistics are listed in tabular format for general interest and applications to Shuttle landing flight rules.

5.1 Summary of Upper-Air Forecast Accuracy

Meso-Eta error characteristics for upper-air forecasts at XMR, TBW, and EDW remained generally unchanged between 1996 and 1997 except near the lowest few layers. These characteristics are summarized in Tables 5.1 and 5.2. The model's error growth during the forecast cycle is minimal (section 5.3), implying that all data may be combined into a single dataset regardless of duration. Moreover, since error growth is minimal, the characteristics outlined in Tables 5.1 and 5.2 apply, on average, at any time during the forecast period. This generality does not apply to the surface data where error characteristics varied with time of day. Results are summarized for both XMR and TBW together since their basic characteristics were similar.

Table 5.1. Summary of Meso-Eta upper-air forecast error characteristics at XMR and TBW.

Warm Season (May – Aug)	Cool Season (Oct – Jan)
On average, forecasts are about 1 °C too cold below 700 mb and 1 to 2 °C too warm above 700 mb.	The height of the lower tropospheric temperature inversion is often overforecast, thereby creating a 2 °C cold bias near the 700-mb level.
Forecasts are too dry below 800 mb and too moist above 500 mb.	Wind speed forecasts are about 1 m s ⁻¹ too slow in the middle troposphere and about 1 m s ⁻¹ too fast in the upper troposphere.
The temperature and moisture biases indicate that forecast soundings are too stable on average. This could be a consequence of the model's convective rainfall parameterization.	Wind direction forecast biases are less than ±10°, but the random error component of 10 to 40° dominates the day-to-day variability.
Wind speed forecasts are nearly unbiased in the lower and middle troposphere, but are typically too fast above 400 mb.	
Wind direction forecast biases are less than ±10° but the random error component of 40 to 60° dominates the day-to-day variability.	

Table 5.2. Summary of Meso-Eta upper-air forecast error characteristics at EDW.

Warm Season (May – Aug)	Cool Season (Oct – Jan)
Temperature biases are less than ± 1 °C.	A strong cold bias exists in the forecasts below 700 mb. The bias exceeds -4 °C near the surface.
Forecasts tend to retain greater amounts of moisture than observed except near the 600 mb level.	Forecasts are too moist near the surface, and too dry above 800 mb.
Wind speed forecasts are 1 to 2 m s^{-1} too slow, but the random error component of 3 to 5 m s^{-1} dominates the day-to-day variability.	Wind speeds are 1 to 2 m s^{-1} too slow on average except near the tropopause. The random error component exceeds 6 m s^{-1} .
On average, wind direction forecasts are backed about 10° relative to observations. The random error component of 30 to 60° dominates the day-to-day variability.	On average, wind direction forecasts are backed about 10° relative to observations. The random error component of 30 to 90° dominates the day-to-day variability.

5.2 Detailed Results

The results summarized above in Tables 5.1 and 5.2 are described more completely in this section. The following discussion is useful because it compares the relationship between systematic and nonsystematic error in the forecasts. Moreover, speculations are offered that might explain possible sources for errors.

Understanding the source of model error could help forecasters cope with day-to-day variations in upper-air point forecast error. For example, the results below demonstrate that the forecasts at XMR and TBW tend to be more thermodynamically stable than observed. However, in the subjective portion of the Meso-Eta evaluation, MN97 demonstrated that the model typically generates warm season precipitation over Florida too early in the day. One possible explanation for this discrepancy between overly stable forecast soundings and excessive precipitation is that the model erodes the stable surface layer and breaks the capping inversion more quickly and vigorously than observed. Unfortunately, rawinsonde data are not consistently available at sufficient temporal resolution throughout the daytime hours to validate this hypothesis.

5.2.1 Temperature

Warm season temperature biases at EDW are less than ± 1 °C (Fig. 5.1a). At XMR and TBW, forecast temperatures below 700 mb are about 1 °C colder than observed whereas above 700 mb they are about 1 to 2 °C warmer than observed. The net effect for warm season forecasts at the Florida stations is a tendency towards a thermally stable model atmosphere. RMS errors range from about 1 to 2.5 °C and are largest in the upper troposphere (Fig. 5.1b). In comparison, typical RMS uncertainty in rawinsonde temperature observations is about 0.6 °C (Hoehne 1980; Ahnert 1991). This fact suggests that about half of the nonsystematic error between the forecasts and observations may be due to measurement uncertainty.

During the cool season, temperature forecasts at EDW exhibit a negative (cold) bias below 700 mb that exceeds -4 °C near the surface (Fig. 5.1d). At XMR and TBW, temperature biases are less than 1 °C except around the 700 mb level and above the tropopause. Examination of individual forecast and observed soundings at XMR throughout the cool season (not shown) reveals that the 700 mb cold bias appears primarily because model forecasts of the lower tropospheric inversion height are frequently at a higher level than where they are actually observed. In the middle troposphere, RMS errors in cool season temperature forecasts at EDW are substantially larger those at XMR and TBW (Fig. 5.1e). Since biases are small above 700 mb at EDW, the relatively large error standard deviations suggest that a greater portion of the total RMS error is caused by a large amount of day-to-day variability in the forecast errors (Fig. 5.1f).

5.2.2 Mixing Ratio

Warm season mixing ratio biases at XMR and TBW (Fig. 5.2a) indicate that Meso-Eta forecasts are on average about 1 g kg^{-1} too dry below 700 mb. Conversely, mixing ratio biases at EDW are about 0.5 g kg^{-1} greater than observed. Between 700 and 500 mb, forecasts at all three locations indicate a negative (dry) bias while above 500 mb they tend to retain excessive amounts of moisture. In combination with the negative lower tropospheric temperature biases discussed previously, these results suggest that warm season model forecasts at XMR and TBW are typically more stable than observed. Cool season mixing ratio biases at all three locations reveal excessive moisture near the surface with a rapid vertical transition to a layer with less moisture than observed (Fig. 5.2d).

RMS errors for the warm season (Fig. 5.2b) drop from around 2.5 g kg^{-1} at low-levels (1.5 g kg^{-1} at EDW) to near zero at 200 mb, where there is very little water vapor present in the atmosphere. In the cool season, RMS errors follow a similar profile at all three stations starting with values of 2 g kg^{-1} near the surface (Fig. 5.2e). Since the error standard deviations shown in Figs. 5.2c and 5.2f are more than double the magnitude of the mixing ratio biases, nonsystematic errors account for roughly 50 to 75% of the total RMS error. Results shown in Figs. 5.2b and 5.2e are consistent with those of Rogers et. al (1996), who show 24-h RMS errors in specific humidity from 48-km Eta model forecasts across the United States during September 1994 ranging from nearly 2 g kg^{-1} at 1000 mb to less than 0.1 g kg^{-1} at 250 mb (see their Fig. 7). Note that these calculations for mixing ratio errors are not normalized by magnitude and are therefore not representative of percent errors as the mixing ratio tends toward zero in the upper troposphere.

5.2.3 Wind Speed

Warm season wind speed biases are generally less than $\pm 1 \text{ m s}^{-1}$ (Fig. 5.3a). The exception occurs at EDW where lower tropospheric wind speed forecasts are about 2 m s^{-1} slower than observed. This result is consistent with the negative (slow) bias in 10-m wind speed forecasts identified at EDW (Fig. 4.1j). Below 400 mb, warm season RMS errors range from about 2 to 4 m s^{-1} (Fig. 5.3b). RMS errors around the 200 mb level are larger with values approaching 6 m s^{-1} . For comparison, uncertainties in rawinsonde wind speed measurements are about 3.1 m s^{-1} (Hoehne 1980). The relatively large nonsystematic error component is derived from some combination of measurement uncertainty and insufficient model resolution.

During the cool season, forecast wind speeds at XMR and TBW are about 1 m s^{-1} slower (faster) than observed in the middle (upper) troposphere (Fig. 5.3d). At EDW, wind speed biases range from 1 to 3 m s^{-1} except near the surface where forecast wind speeds remain slow. Cool season RMS errors at XMR and TBW are comparable to those found during the warm season and again, receive large contributions from the nonsystematic error component (Fig. 5.3e). At EDW, cool season RMS errors above 700 mb are nearly double those of the warm season with increased contributions from both systematic and nonsystematic errors.

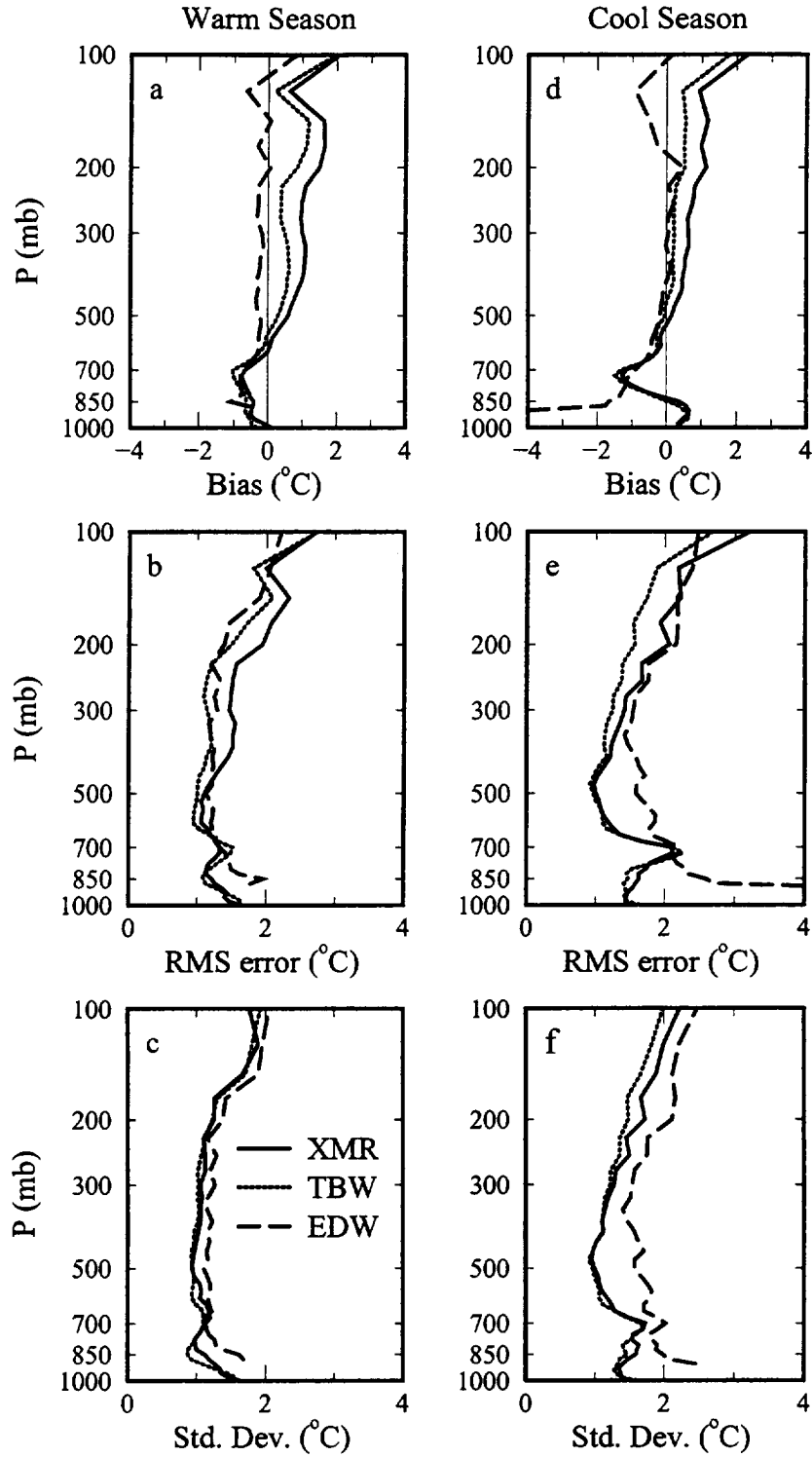


Figure 5.1. Bias (forecast – observed), RMS error, and error standard deviation (°C) of temperature forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

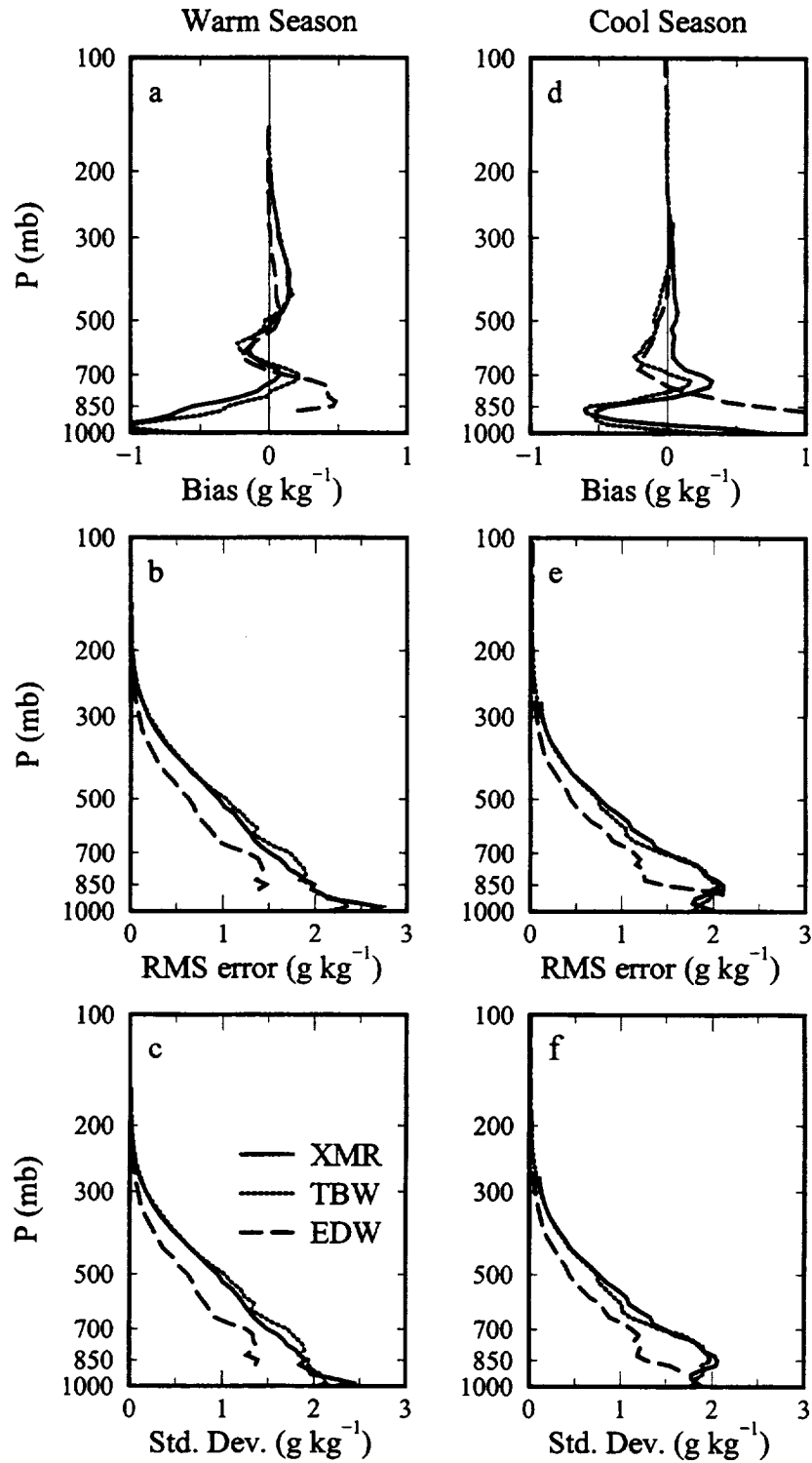


Figure 5.2. Bias (forecast – observed), RMS error, and error standard deviation (g kg^{-1}) of mixing ratio forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

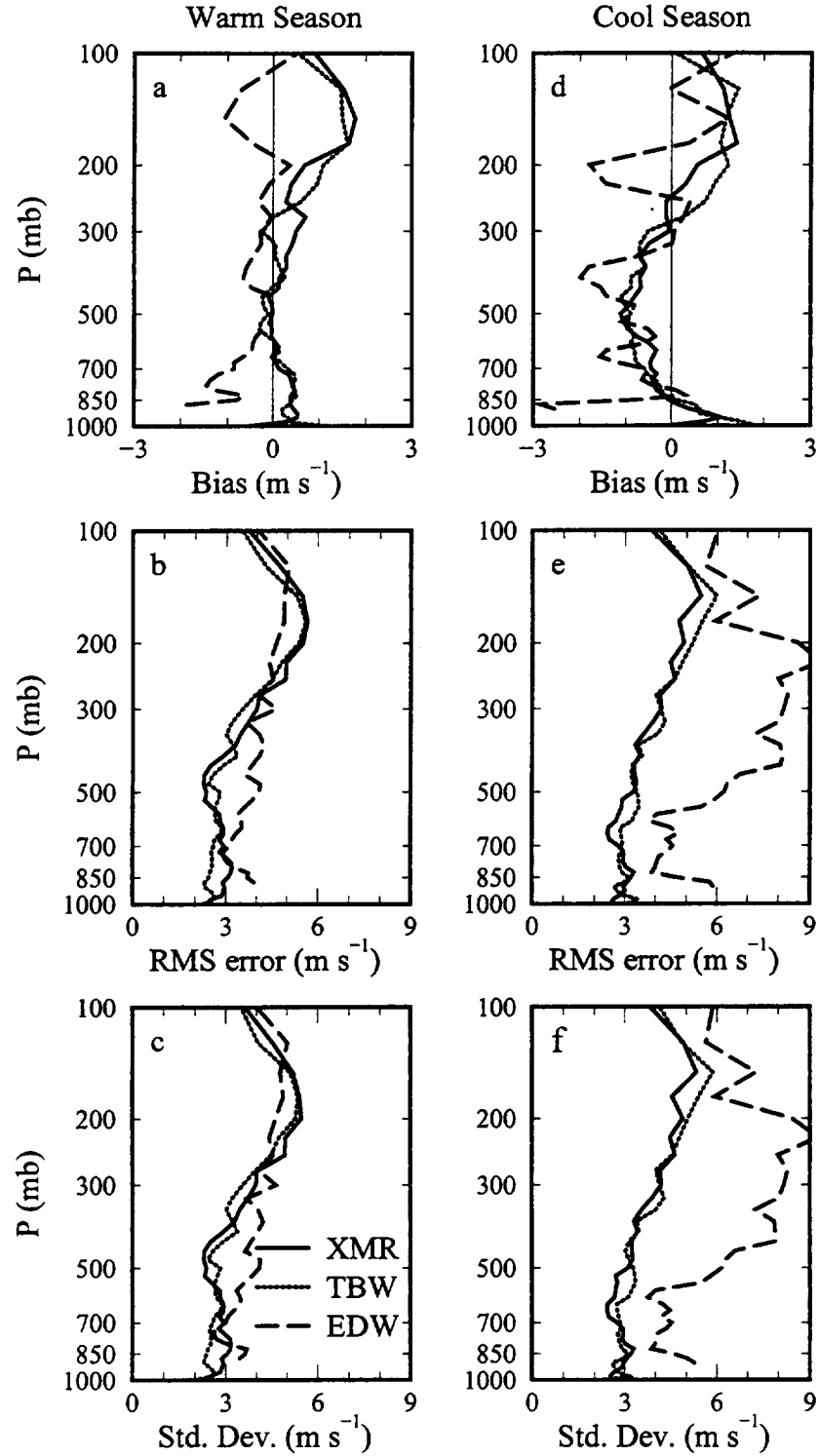


Figure 5.3. Bias (forecast – observed), RMS error, and error standard deviation (m s^{-1}) of wind speed forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

5.2.4 Wind Direction

Warm season wind direction forecasts at XMR and TBW are nearly unbiased, as mean errors lie within $\pm 10^\circ$ (Figs. 5.4a). Biases at EDW are within $\pm 20^\circ$ and demonstrate greater variability with height, especially within the lower troposphere. Moreover, warm season wind direction forecasts at EDW are typically negative, or counter-clockwise relative to observed values. Although forecast wind direction biases are reasonably small, RMS errors and standard deviations approach 60° and are nearly twice as large at XMR and TBW as they are at EDW in the upper troposphere (Figs. 5.4b, c). Hoehne (1980) reports that observational uncertainty in wind direction decreases with increasing wind speeds from 14° at 10 kt to 2° at 120 kt. In addition to the fact that observational uncertainty for wind direction is a function of wind speed, the accuracy of forecast wind directions may also be influenced by wind speed. At most levels, average warm season observed wind speeds (not shown) at EDW are about 5 to 15 kt faster than those at XMR and TBW with larger differences occurring in the upper troposphere. Average warm season forecast wind speeds exhibit the same characteristics. Since mean forecast and observed wind speeds are faster at EDW, it is not surprising that standard deviations in wind direction are smaller at EDW than those at XMR and TBW. That is, the difference in wind direction standard deviations between EDW and XMR and TBW could be accounted for, in part, by difficulties in both observing and forecasting wind directions accurately at low wind speeds.

Cool season errors in wind direction forecasts are similar to those of the warm season with a few exceptions. RMS errors and standard deviations at XMR and TBW decrease relative to warm season values (Figs. 5.4b, c, e, f). During the cool season, both forecast and observed average wind speeds (not shown) increase relative to warm season speeds by about 15 to 25 kt at XMR and TBW. Therefore, the upper tropospheric decrease in RMS errors and standard deviations in wind direction at XMR and TBW could again be explained by the idea that both forecast and observed wind directions are more accurate at higher wind speeds.

5.2.5 Geopotential Height

Forecast and observed geopotential heights are computed at 25-mb increments using the hypsometric equation (Wallace and Hobbs 1977, pg. 57). The thickness is determined for each 25-mb layer from the surface to 100 mb and subsequently added to form an integrated height profile. By definition, the thickness of each layer is proportional to its mean virtual temperature. Forecast errors in virtual temperature are therefore manifested as an integrated error in geopotential height that becomes larger with height (Fig. 5.5). Since the difference between temperature and virtual temperature is typically less than one percent, geopotential height errors (Fig. 5.5) correspond qualitatively with the upper-level temperature errors discussed earlier (Fig. 5.1).

Geopotential height forecasts are nearly unbiased at EDW during the warm season (Fig. 5.5a) but cool-season errors exhibit substantial day-to-day variability (Fig. 5.5f). The relatively large height bias over the middle- and upper-troposphere at TBW and XMR is consistent with the integrated effect of the positive (warm) bias in temperature (Fig. 5.1).

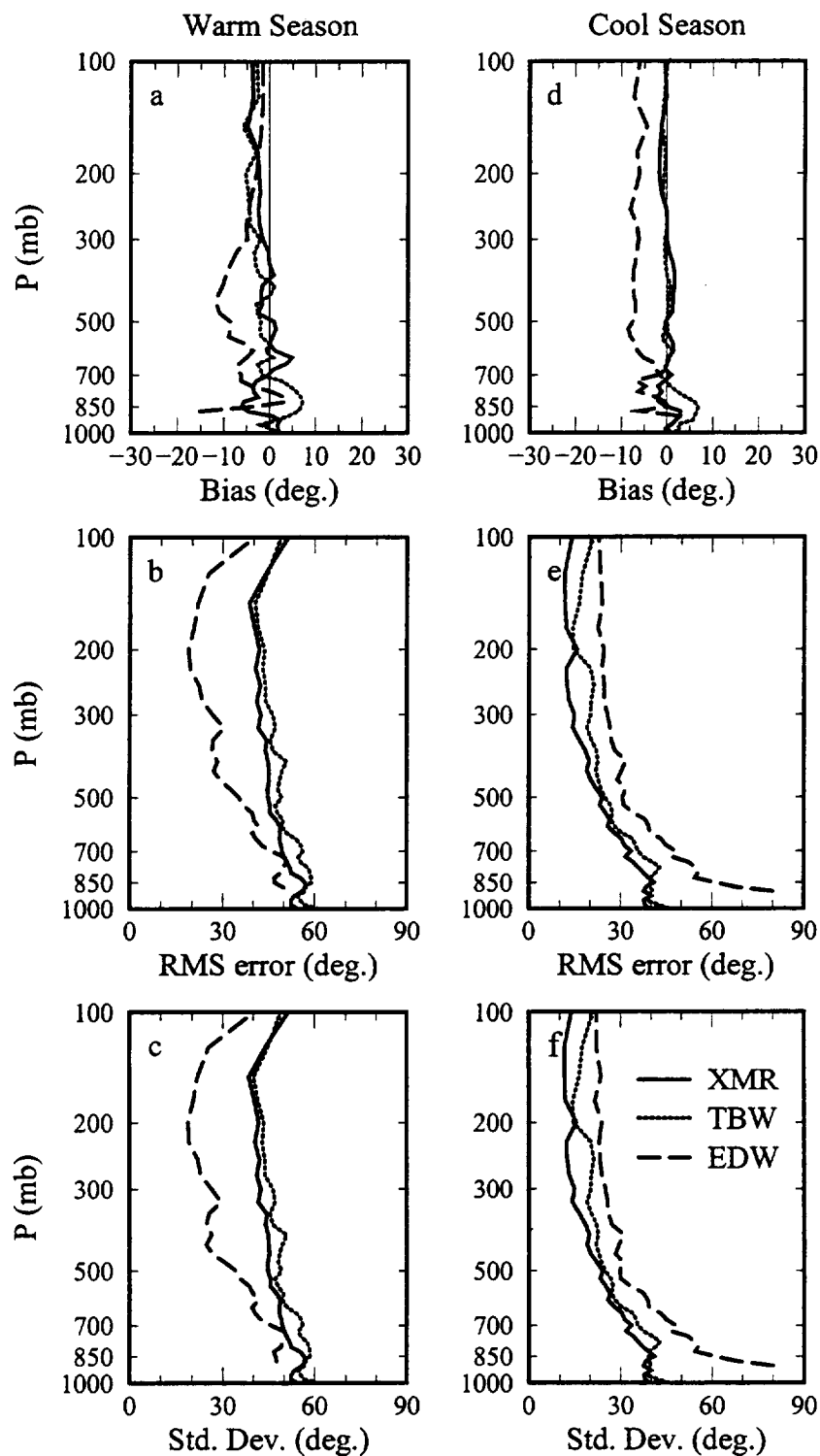


Figure 5.4. Bias (forecast – observed), RMS error, and error standard deviation ($^{\circ}$) of wind direction forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

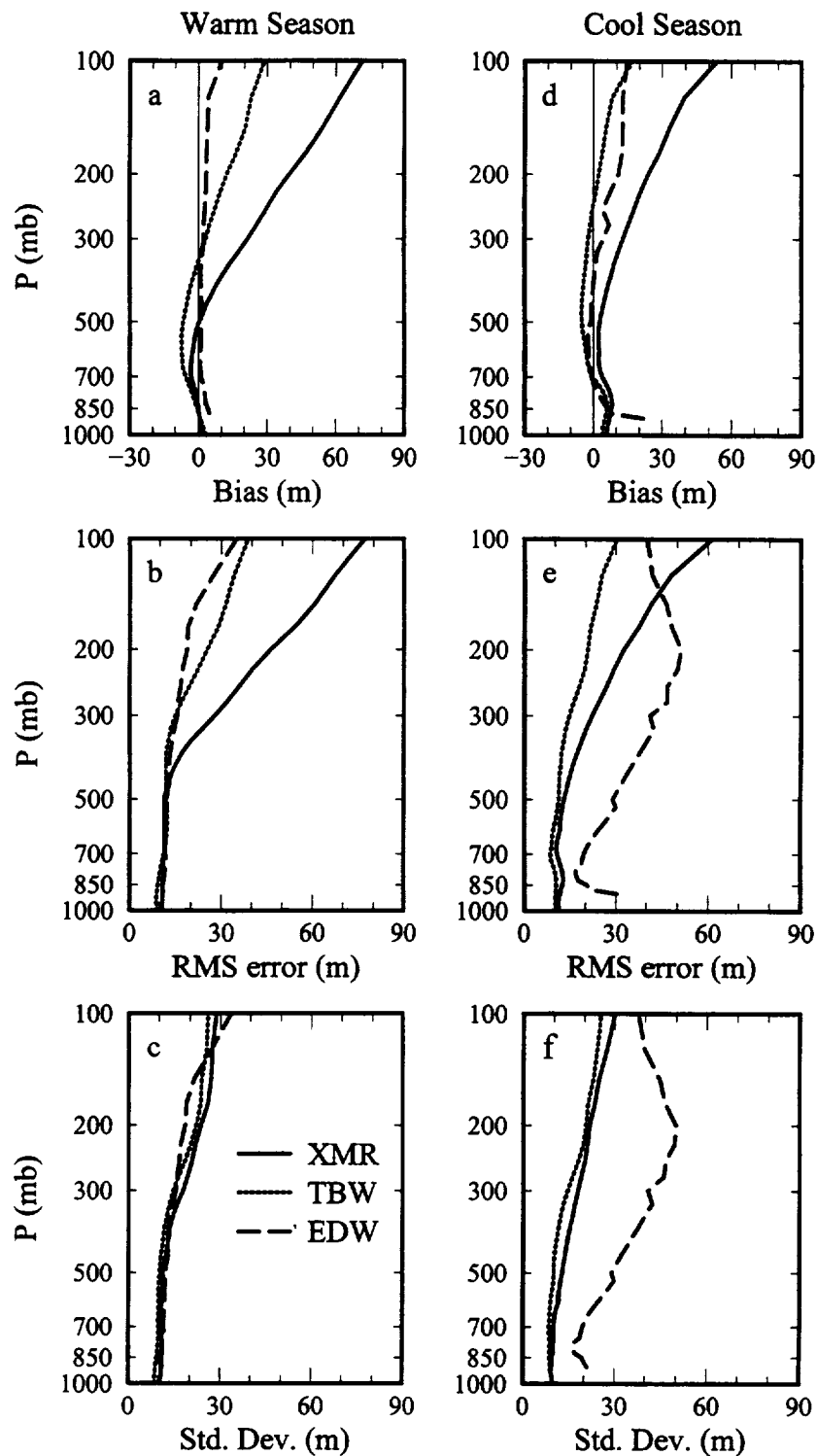


Figure 5.5. Bias (forecast – observed), RMS error, and error standard deviation (m) of geopotential height forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

5.3 Forecast Error Growth

In sections 5.1 and 5.2, it was stated that the Meso-Eta model's systematic error growth during the forecast cycle is minimal. Following this assumption, all data were combined into a single dataset for verification regardless of forecast duration. This statement is validated here. The result is valuable for operational forecasters since the error characteristics outlined in Tables 5.1 and 5.2 apply, on average, at any time during the forecast cycle.

Since rawinsonde observations are available only twice daily under normal circumstances, it is not possible to observe the temporal evolution of upper level forecast errors on an hourly basis throughout the forecast cycle¹. However, separate examination of seasonal forecast errors at three 12-h intervals reveals that upper-level errors do fluctuate slightly with forecast duration although their vertical profiles remain qualitatively similar (MN97). Unlike the surface error characteristics, diurnal oscillations are not evident in the upper-air forecast errors above the lowest few levels. A paired Z statistic is therefore used to determine if seasonal mean changes in upper-level model biases during a 24-h period represent a statistically significant systematic error growth (see Appendix B).

Examination of statistics at each of the three 12-h verification intervals (not shown) reveals that the lower tropospheric cold bias in forecast temperatures (e.g. Fig 5.1) at XMR and TBW becomes more negative with time during both warm and cool seasons. The corresponding paired Z statistic (Figs. 5.6a, d) indicates that this growing cold bias is statistically significant over a 24-h period. Following a similar argument, the positive bias in upper tropospheric temperature forecasts during the warm season at TBW tends to grow stronger with time (Fig. 5.6a). The significance of the warm season error growth in forecast temperatures near the surface at EDW (Fig. 5.6a) is questionable due to the possible influence of diurnal variability (e.g., Fig. 4.1a). Most error growth in mixing ratio forecasts is not statistically significant except near the 200 mb level at TBW (Figs. 5.6b, e). The only significant change in systematic error (bias) for wind speed forecasts is found near the tropopause during the warm season at XMR and TBW (Fig. 5.6c). Examination of statistics at each of the three 12-h verification intervals (not shown) reveals that the positive (fast) bias in upper tropospheric wind speed forecasts at these locations (e.g. Fig 5.5a) tends to diminish with time during the warm season.

Although a few exceptions are noted here, the results shown in Fig. 5.6 reveal that the 24-h systematic error growth for temperature, mixing ratio, and wind speed is not statistically significant at the 99% confidence level. For the purpose of establishing generalized forecast error characteristics, this result validates our assumption that sample sizes can be maximized by combining all available upper-air profiles.

¹ The 50 MHz wind profiler data at KSC/CCAS are available every 5 min but are not used for the objective portion of this study because similar data are not available at TBW or EDW.

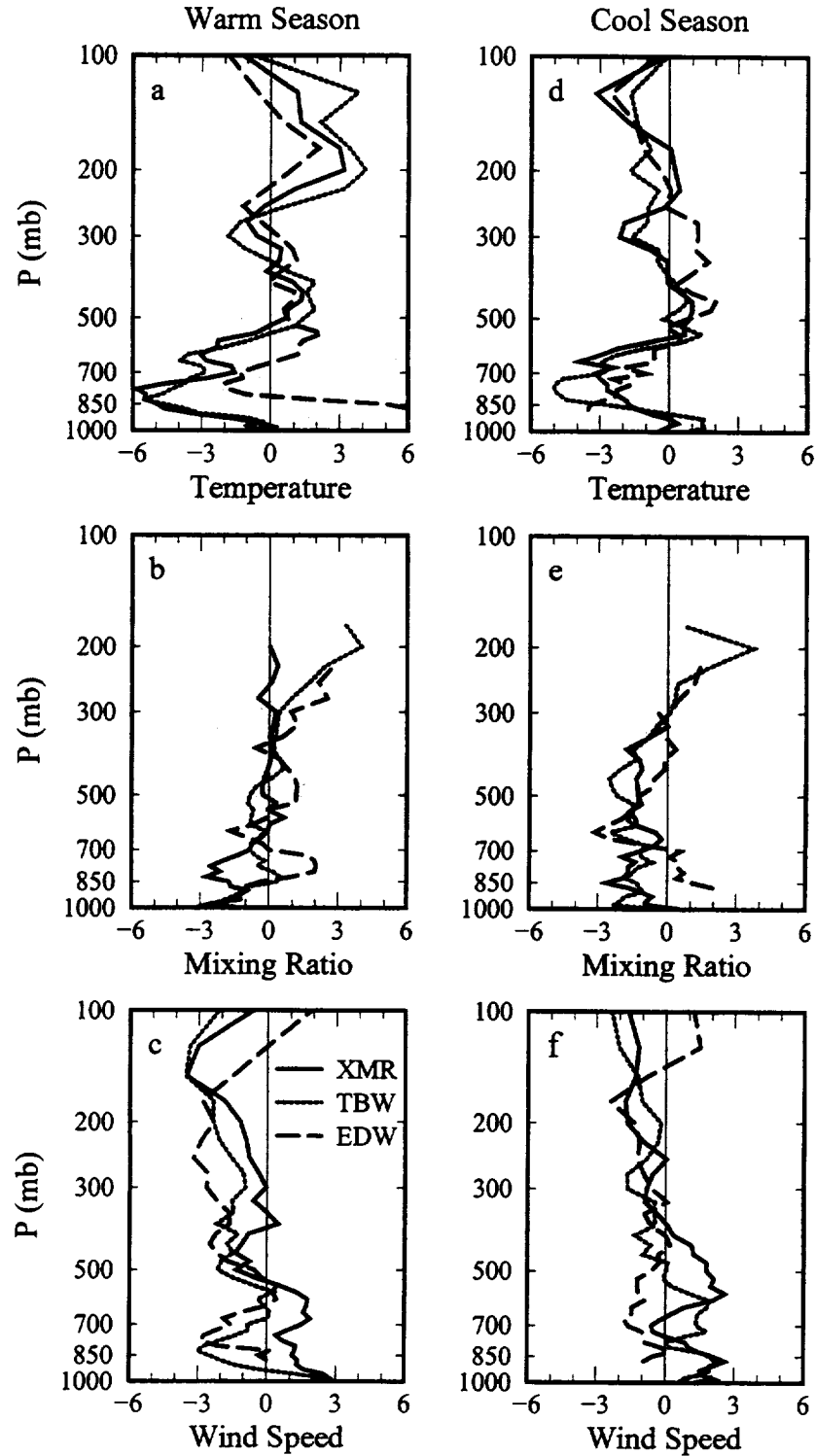


Figure 5.6. Paired Z statistic plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). The nondimensional statistic is shown for temperature (a, d), mixing ratio (b, e) and wind speed (c, f). Warm season values are shown in the left column (panels a-c) while cool season values are shown in the right column (panels d-f). Paired Z values that lie outside the shaded region indicate that 24-h systematic error growth is statistically significant at the 99% confidence level (see Appendix B).

6.0 Convective Indices

Forecast errors are examined for convective indices specified in Table 1.1. Results are limited to XMR and TBW since convection is usually not a concern at EDW. Except for the 0-3 km helicity (Lilly 1986; Davies-Jones et al. 1990) and MDPI (Wheeler and Roeder 1996), convective indices are computed using the numerical algorithms described in the GEneral Meteorological PAcKage (GEMPAK; desJardins et al. 1997). As with the upper-air forecast verification, the convective index error statistics are computed using all available forecast/observation pairs during 1996 and 1997 from both the 0300 UTC and 1500 UTC model runs.

Generalized results of the verification are summarized in section 6.1. However, a more complete analysis of convective index error characteristics is provided in section 6.2. Forecast errors are quantified using traditional bias (forecast – observed), RMS error, and error standard deviation statistics (Tables 6.2, 6.3). The characteristics of the forecasts, observations, and their relationship are also described more completely using *conditional quartile diagrams* (Murphy et al. 1989).

The conditional quartile diagram helps visualize the relationship between given forecasts and observations and can aid operational users by serving as a lookup diagram based on past data. For example, given a particular forecast value for a convective index, the user can determine its bias, the expected range of observed values, and its frequency of occurrence relative to other forecast values. A rigorous explanation of conditional quartile diagrams is provided in Appendix D, but its basic features are as follows.

- A frequency histogram (thin vertical lines) describes the distribution of all Meso-Eta forecasts verifying at the rawinsonde observation times. This histogram helps the forecaster determine if a given forecast is relatively common, or if it might be considered an extreme event.
- At each of the forecast intervals indicated along the *x*-axis by the frequency histogram, the verifying observations are collected into a *conditional distribution of observations*. Each of these distributions of observations are *conditional* based on the occurrence of a particular forecast value. The conditional distribution of observations helps forecasters determine the range of observations that are expected given a particular forecast value.
- The conditional distribution of observations is summarized for each given forecast value by the median, interquartile range (IQR), and minimum and maximum values. The median represents the middle value of a set of ordered data and is alternately called the 50th percentile or 2nd quartile. Similarly, the IQR represents the difference between the 3rd and 1st quartiles (75th and 25th percentiles). Again, these quantities help forecasters determine the range of observations that are expected given a particular forecast value.
- A 45° reference line is provided to help indicate conditional biases in the forecasts. If the median falls below the reference line, the forecast values are more often greater than the observations associated with that forecast. If the median lies above the reference line, the forecast values are more often less than the observations associated with that forecast.

6.1 Summary of Convective Index Error Characteristics

Convective indices are vertically integrated quantities. Therefore, slight errors in the forecast vertical profiles of temperature, moisture, and/or winds are exacerbated through the computation of convective index error characteristics. The day-to-day variability in observed convective indices is not predicted well by the Meso-Eta model. The errors tend to be smallest during the cool season when, under normal circumstances, they provide little added value for operational forecasting purposes. Some of the more generalized warm season results are summarized in Table 6.1.

Table 6.1. Summary of Meso-Eta point forecast error characteristics for convective indices at XMR during the 1996 and 1997 warm seasons (May-Aug).

Index	Forecast Error Characteristics
Precipitable Water	Forecasts have a slight dry bias, but are generally accurate across a wide range of values.
Lifted Index	Forecasts are more stable than observed. The forecasts tend to be most accurate when their values are around -3 to -4 °C but the day-to-day variations are not handled well.
K-index	Forecast biases are small but a large random error component limits their utility.
LCL	Forecast accuracy decreases at lower pressures (greater heights).
CAPE	Forecasts are too small (stable) on average and are susceptible to very large errors.
0-3 km Helicity	Forecasts tend to overestimate the magnitude of the vertically integrated wind shear.
MDPI	Forecasts are most reliable when values are near 1.0 but a large random error component limits their utility.

6.2 Convective Index Error Characteristics

6.2.1 Precipitable Water

The frequency distribution of forecast PWAT is comparable during the warm season at both XMR and TBW (Figs. 6.1a, b). A slight conditional bias indicates that the observed PWAT is often underforecast by about 0 to 2 mm. This characteristic is confirmed in a more general sense by the overall negative (dry) forecast bias during the warm season (Table 6.2). The IQR varies from about 5 to 10 mm while the conditional minimum and maximum observations are distributed evenly relative to the given forecast values. Although the histogram of forecast PWAT is broader during the cool season, the character of the relationship between the forecasts and their observations remains mostly unchanged except for a reduction of the conditional dry bias (compare Figs 6.1a-d and Tables 6.2 and 6.3). In general, forecasts for PWAT are reliable and well refined.

6.2.2 Lifted Index

Warm season lifted index forecasts are distributed from about -5 to 4 °C (Fig. 6.2a, b). A conditional bias is evident which suggests the forecast soundings typically have greater thermal stability than observed (i.e., forecast lifted index is greater than observed, on average). This positive bias appears in the overall sample statistics (Table 6.2) and was also identified previously in the form of a warm bias in the temperature error profile (Fig. 5.1a). Notably, the conditional bias is smallest when the model predicts lifted indices around -3 to -4 °C. In combination with the overall stable bias, the large variability indicated by the conditional IQR and extreme observations suggest that day-to-day fluctuations in lifted index are not well represented by the Meso-Eta model throughout the warm season. During the cool season, the forecast lifted indices are distributed across a wider range of values (Figs. 6.2c, d). Since the slope of the cool season conditional median is close to that of the 45° reference line, the conditional bias remains small throughout nearly the entire given range of forecast values. Clearly, lifted index forecasts are most reliable during the cool season.

Table 6.2. Bias, RMS error and error standard deviation for warm season (May - August) convective parameters at XMR and TBW.

Convective Index	Bias		RMS Error		Std. Dev.	
	XMR	TBW	XMR	TBW	XMR	TBW
Precipitable Water	-1.81	-1.49	5.20	5.54	4.87	5.34
Lifted Index	1.90	1.44	3.22	2.81	2.61	2.41
K-Index	-.37	.97	6.79	7.07	6.78	7.00
LCL	-5.63	-1.96	32.72	35.46	32.23	35.41
CAPE	-876.91	-543.34	1461.72	1104.69	1169.46	961.84
Convective Inhibition	-1.34	-9.31	65.75	63.56	65.74	62.88
Helicity	5.21	3.58	40.46	37.89	40.12	37.72
MDPI	-.18	-.15	.35	.31	.31	.27

Table 6.3. Bias, RMS error and error standard deviation for cool season (October - January) convective parameters at XMR and TBW.

Convective Index	Bias		RMS Error		Std. Dev.	
	XMR	TBW	XMR	TBW	XMR	TBW
Precipitable Water	.31	-1.37	4.84	5.01	4.83	4.82
Lifted Index	-.61	-.06	2.90	2.83	2.83	2.83
K-Index	2.19	-1.24	12.11	11.29	11.91	11.22
LCL	16.58	10.64	39.53	36.65	35.88	35.07
CAPE	19.88	7.13	393.51	291.89	393.01	291.81
Convective Inhibition	1.88	2.21	32.55	47.05	32.50	46.99
Helicity	10.93	4.17	77.21	79.69	76.43	79.58
MDPI	.06	.05	.27	.24	.26	.24

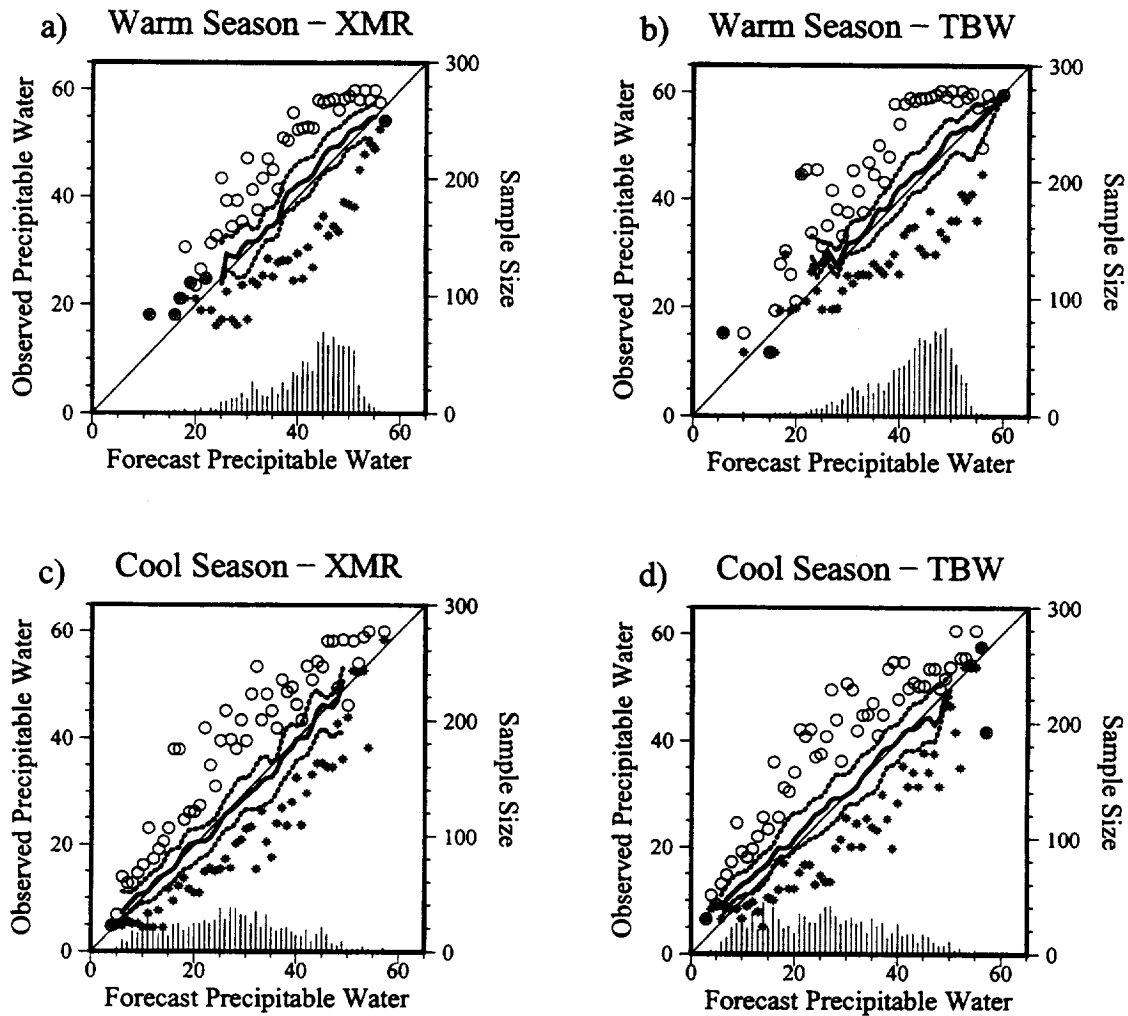


Figure 6.1. Quartile diagram showing the conditional distribution of observed PWAT (mm) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of PWAT (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

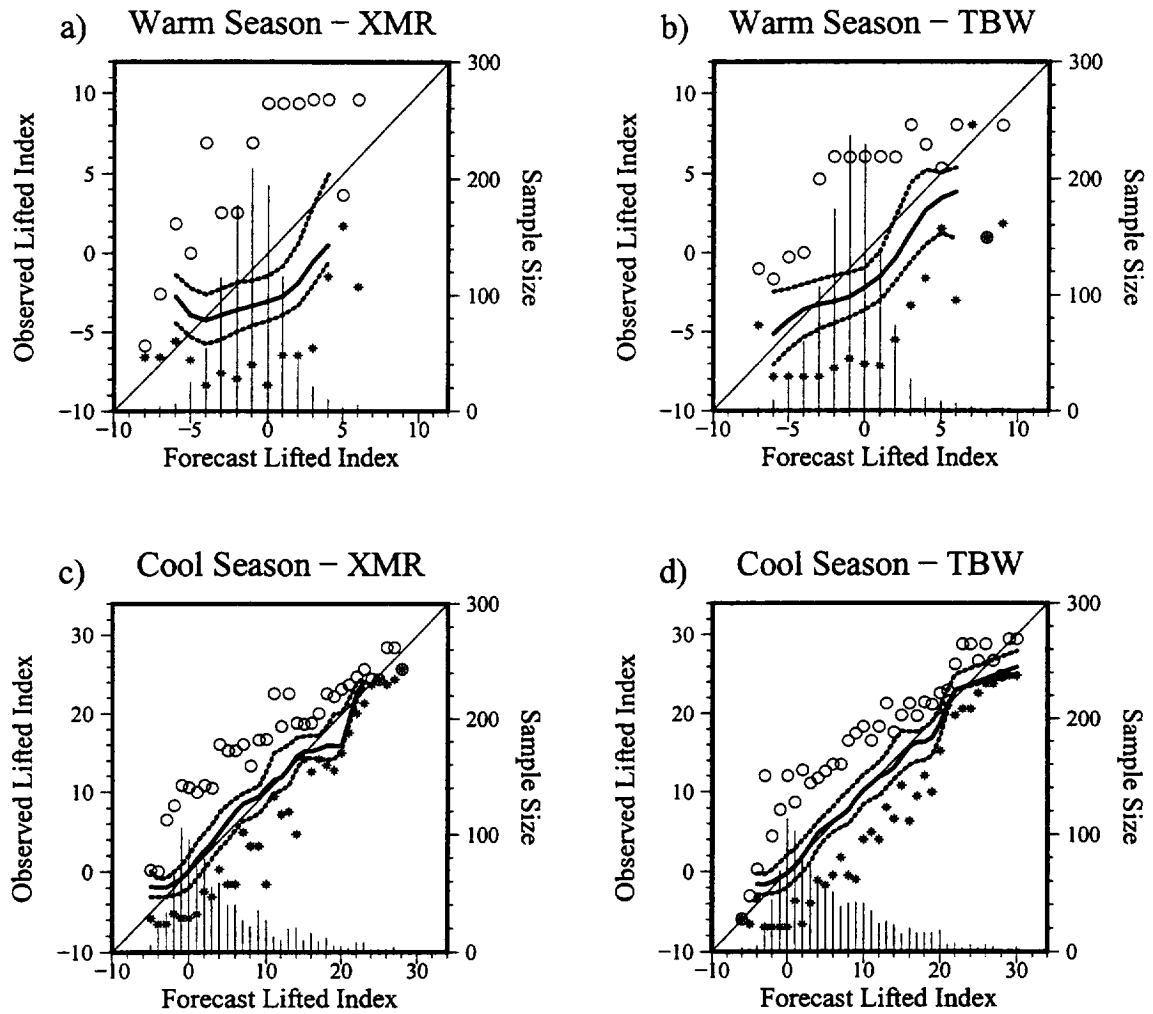


Figure 6.2. Quartile diagram showing the conditional distribution of observed lifted index ($^{\circ}\text{C}$) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of lifted index (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

6.2.3 K-Index

The frequency distribution of warm season K-index forecasts reveals that the indices most often lie in the range from approximately 28 to 32 °C (Figs. 6.3a, b). The conditional bias within this range is small, but the IQR of about 8 °C reveals substantial day-to-day variability in the forecast errors. The overall sample statistics (Table 6.2) also indicate that the K-index forecasts are nearly unbiased and that the error standard deviations are large in comparison to the mean. Cool season K-index forecasts are more widely distributed than the warm season forecasts (Figs. 6.3c, d). When cool season K-index forecasts are negative, they tend to indicate slightly greater stability than observed soundings. Conversely, when positive, the forecasts tend to indicate slightly less stability than observed.

6.2.4 Lifted Condensation Level

When air parcels in the forecast soundings are lifted dry adiabatically, they most frequently reach saturation around 990 to 1000 mb during both warm and cool seasons (Fig. 6.4). The conditional distribution of the observed lifted condensation level (LCL) reveals that forecasts are most reliable when the LCL is closer to the ground. In particular, the large IQR and spread of extreme conditional observations reveals that LCL forecasts are less refined at lower pressures. During the cool season, the conditional median more closely follows the 45° reference line, but the distributions still exhibit a large IQR. Since subtle errors in low-level temperature and moisture will affect the calculation of forecast and observed LCL, it is not surprising that the variability is so large

6.2.5 Convective Available Potential Energy

During the warm season, the forecast convective available potential energy (CAPE) distribution reveals that most forecasts contain nearly zero CAPE and thus are quite stable (Fig. 6.5). When warm season CAPE forecasts are less than about 1500 J kg⁻¹, the conditional median of the observed CAPE distribution is larger than the given forecast values (Figs. 6.5a, b). This result indicates that at smaller values, the warm season CAPE forecasts tend to underestimate the observed instability – a fact which is supported by the negative overall sample bias (Table 6.2). The convective potential of the forecast soundings could be limited by the positive (warm) bias in middle- and upper-tropospheric temperature (Fig. 5.1a) since, in that environment, lifted air parcels would require greater energy to attain positive buoyancy. It is noteworthy that even when zero CAPE is forecast the maximum observed CAPE at XMR may exceed 4100 J kg⁻¹ during the warm season (3000 J kg⁻¹ at TBW). Therefore, warm season CAPE forecasts are susceptible to large errors.

During the cool season, the frequency distribution of forecast CAPE at XMR and TBW reveals that the Meso-Eta model often predicts a relatively stable environment with many forecast soundings again supporting zero CAPE (Figs. 6.5c, d). Unlike warm season results, the cool season forecasts of small CAPE (i.e., near zero) exhibit little conditional bias and thus are reliable. However, the wide distribution of the conditional extreme observations indicates that both warm and cool season CAPE forecasts occasionally suffer large errors. Given cool season forecasts of large CAPE, a conditional bias becomes evident which suggests the environmental instability is overestimated. However, the small sample sizes associated with such forecasts decrease the robustness of this particular result.

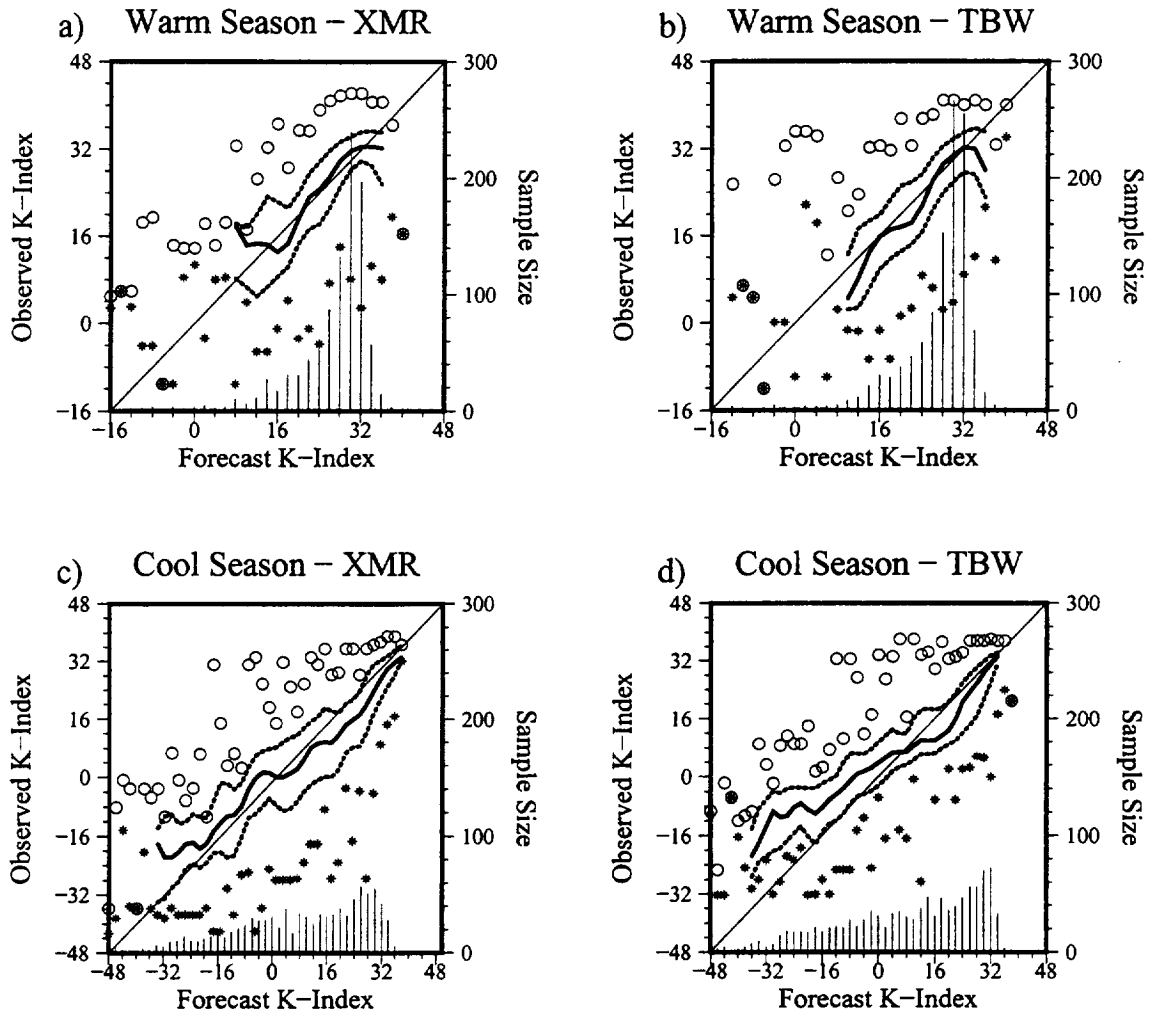


Figure 6.3. Quartile diagram showing the conditional distribution of observed K-index ($^{\circ}\text{C}$) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of K-index (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

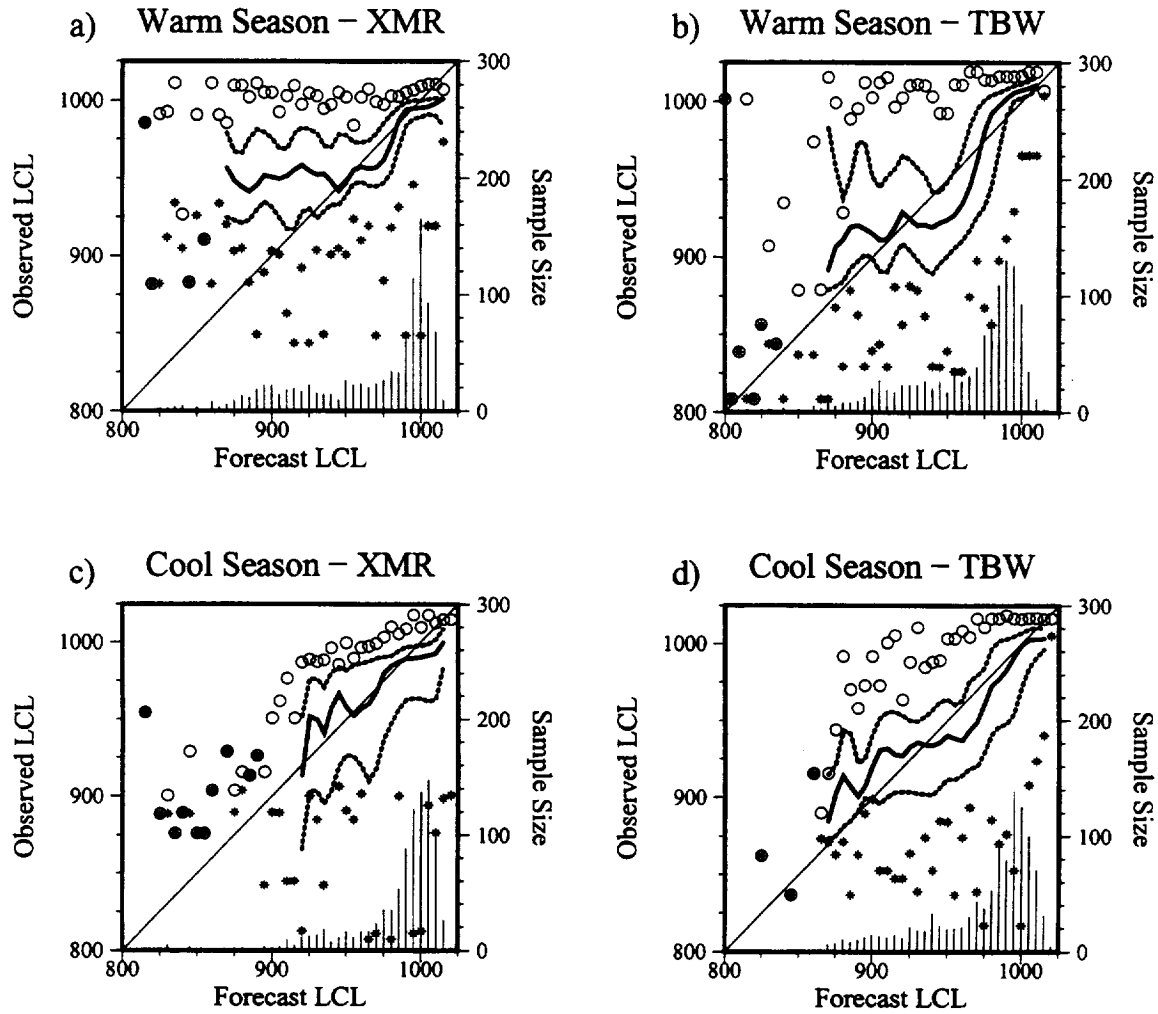


Figure 6.4. Quartile diagram showing the conditional distribution of observed LCL (mb) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of LCL (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

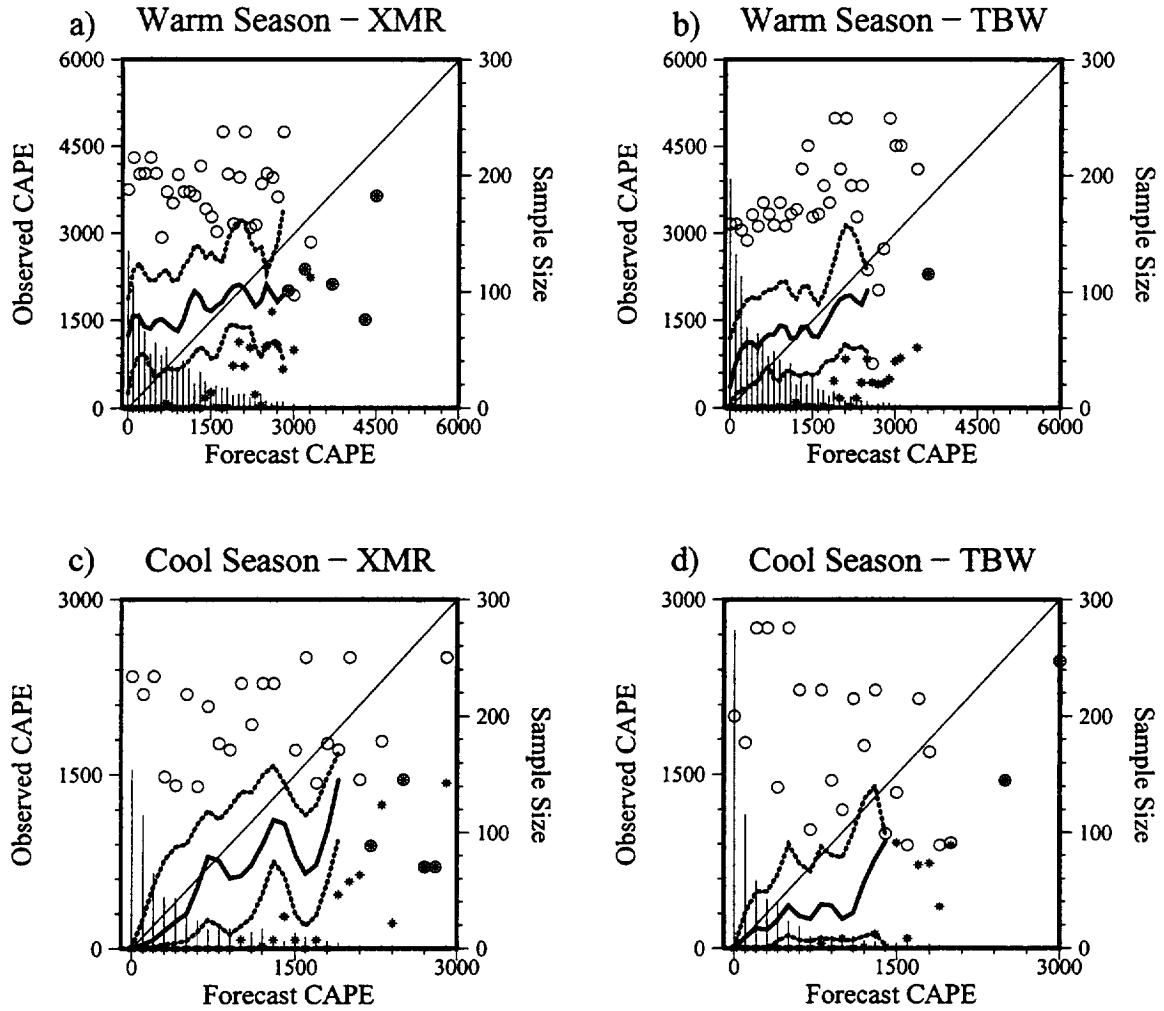


Figure 6.5. Quartile diagram showing the conditional distribution of observed CAPE (J kg^{-1}) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of CAPE (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

6.2.6 Convective Inhibition

The frequency distributions of convective inhibition (CIN; negative area on a skew-T diagram) are similar to those of CAPE forecasts (compare Figs. 6.5 and 6.6). Upon first inspection, the simultaneous and regular occurrence of zero CAPE and CIN may seem in contradiction. However, the explanation is that a closed positive or negative area is not defined on a thermodynamic diagram if a lifted parcel fails to reach a level of free convection or equilibrium level within its environment. Under this circumstance, the GEMPAK numerical algorithm assigns zero CAPE and CIN to the sounding as lifted air parcels remain colder (more stable) than the environmental temperature at all levels. Examination of the warm season data from XMR (not shown) used to construct Figs. 6.3a and 6.4a reveals that instances of both zero CAPE and CIN occur in 4.9 and 2.8% of the available forecasts and observations, respectively. Similarly, 33.9 and 39.3% of the cool season forecasts and observations at XMR exhibit zero CAPE and CIN. The data for TBW demonstrate a comparable relationship. These results indicate that warm season forecasts at XMR and TBW contain zero CIN and CAPE — and are therefore stable — more often than the observed soundings. During the cool season, instances of zero CAPE and CIN occur regularly, though the observed environment is often more stable than the forecast environment.

While considering all available data, the conditional median of the observations reveals that CIN is regularly underforecast during the warm season whenever the forecast CIN is near zero (Figs. 6.6a, b). The fact that warm season forecasts of small (though not necessarily zero) CAPE and CIN both tend to underestimate their corresponding observations suggests that the model's boundary layer error characteristics are different from those in the free atmosphere. Indeed, the vertical profiles of temperature and mixing ratio bias (Figs. 5.1 and 5.2) each exhibit a shift in sign around the 700 mb level. During the cool season, the CIN forecasts are conditionally unbiased when near zero (Figs. 6.6c, d). This cool season reliability likely follows from the frequent occurrence of forecast and observed soundings with zero CAPE and CIN as discussed above. Given larger CIN forecasts, the position of the conditional median relative to the 45° reference line indicates that CIN is usually overforecast during both the warm and cool seasons. However, since forecasts of large CIN are relatively uncommon, this conditional bias should be considered cautiously.

6.2.7 Helicity

Positive and negative 0-3 km helicity (Lilly 1986; Davies-Jones et al. 1990) forecasts are nearly evenly distributed about zero and are thus well refined during both warm and cool seasons at XMR and TBW (Fig. 6.7). Helicity forecasts with values near zero are conditionally unbiased. However, the lower slope of the conditional median relative to the 45° reference line indicates that the magnitude of the helicity is often overforecast, particularly during the warm season. The overall sample biases (Tables 6.2 and 6.3) are small but do not fully represent the tendency to overforecast the magnitude of nonzero helicity values as revealed by Fig. 6.7. The frequency distribution of cool season helicity forecasts covers nearly double the range of values that are forecast during the warm season. The corresponding RMS errors and error standard deviations are also approximately doubled during the cool season (Tables 6.2 and 6.3). Since helicity is a measure of the vertically integrated wind shear over the lowest 3 km of the atmosphere, the larger errors during the cool season are likely related to an increase in forecast and observed wind speeds and corresponding wind shear.

6.2.8 Microburst Day Potential Index

The frequency distribution for Microburst Day Potential Index (MDPI; Wheeler and Roeder 1996) forecasts indicates that most values occur in the range 0.8 to 0.9 during the warm season (Figs. 6.8a, b). The MDPI forecasts are conditionally unbiased and therefore reliable when near 1.0. Given MDPI forecasts less than 1.0, the potential for microburst development is often underforecast relative to the conditional distribution of observed MDPI. Conversely, the microburst potential is typically overestimated as the forecast MDPI exceeds 1.0. Note that at XMR, the conditional maximum observed MDPI reach 1.25 across nearly the entire range of given forecast values (Fig 6.8a). Since MDPI values above 1.0 support a high probability of wet microburst development at XMR (Wheeler and Roeder 1996), this result suggests that the model often fails to accurately predict the environment conducive for observed microbursts. The slight negative bias in the overall sample statistics (Table 6.2) supports the tendency to underestimate observed MDPI. During the cool season, the MDPI forecasts cover a wider range of values (more refined) while the slope of the conditional median more closely matches that of the 45° reference line (more reliable).

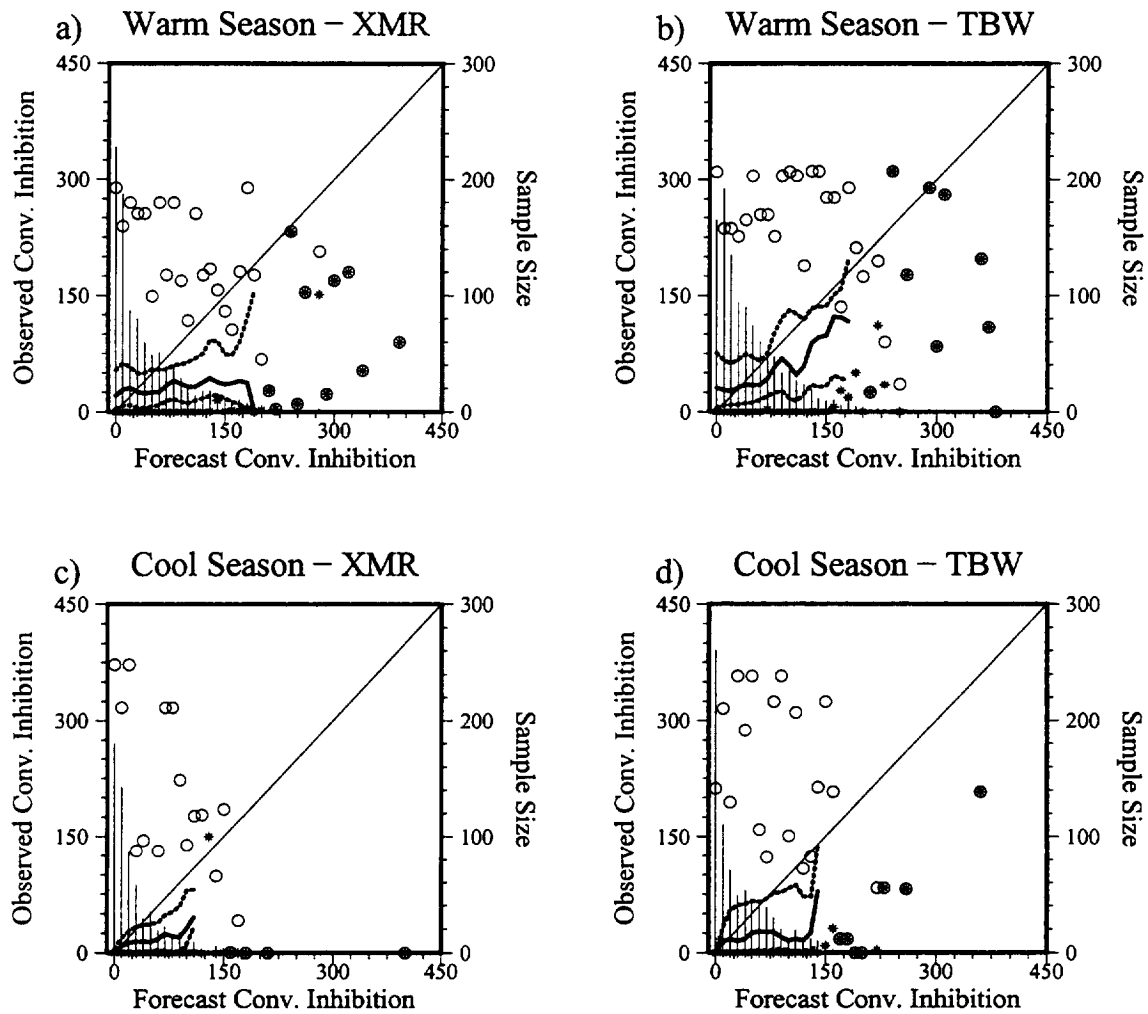


Figure 6.6. Quartile diagram showing the conditional distribution of observed CIN (J kg^{-1}) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of CIN (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

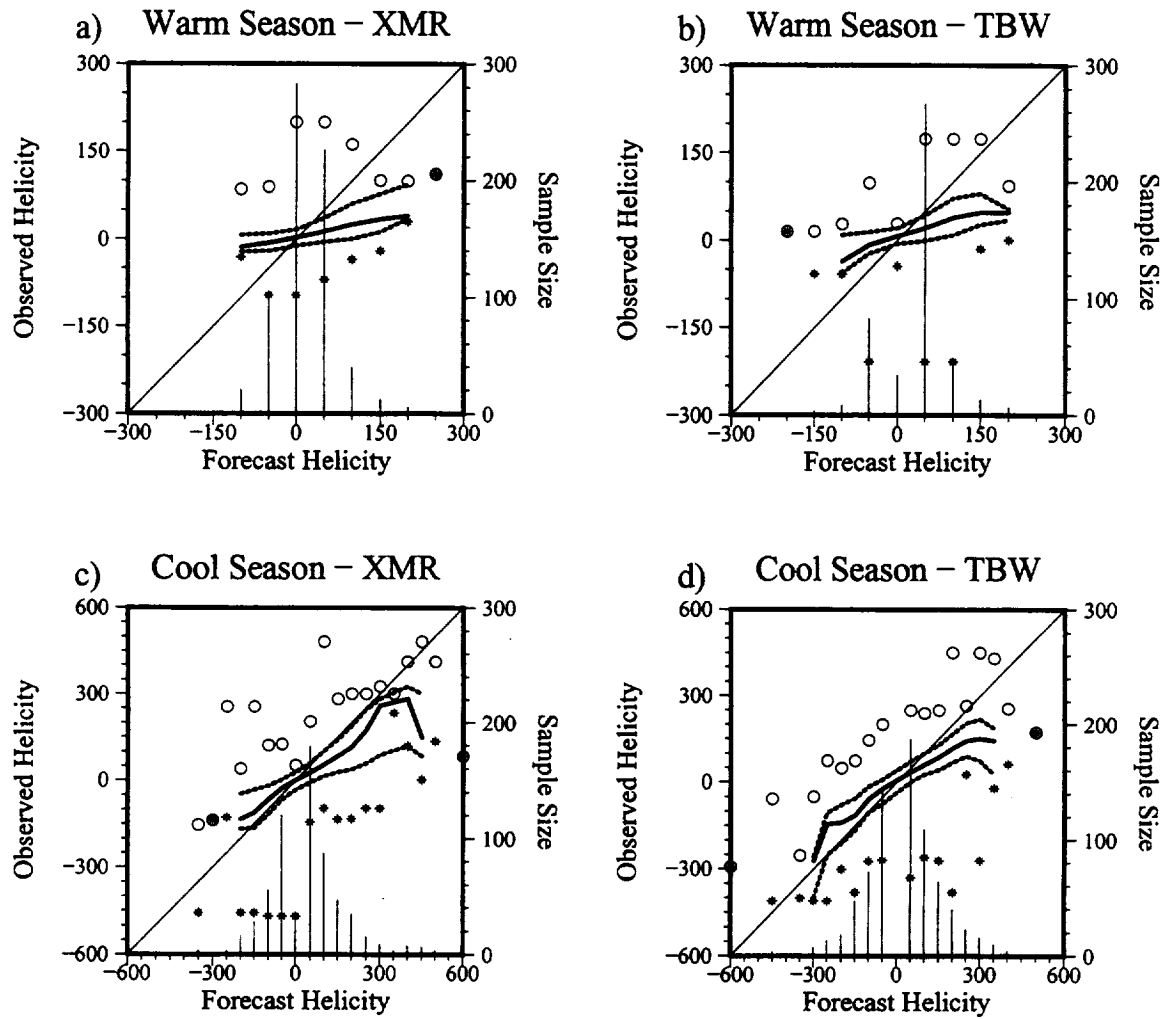


Figure 6.7. Quartile diagram showing the conditional distribution of observed helicity ($\text{m}^2 \text{s}^{-2}$) associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of helicity (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

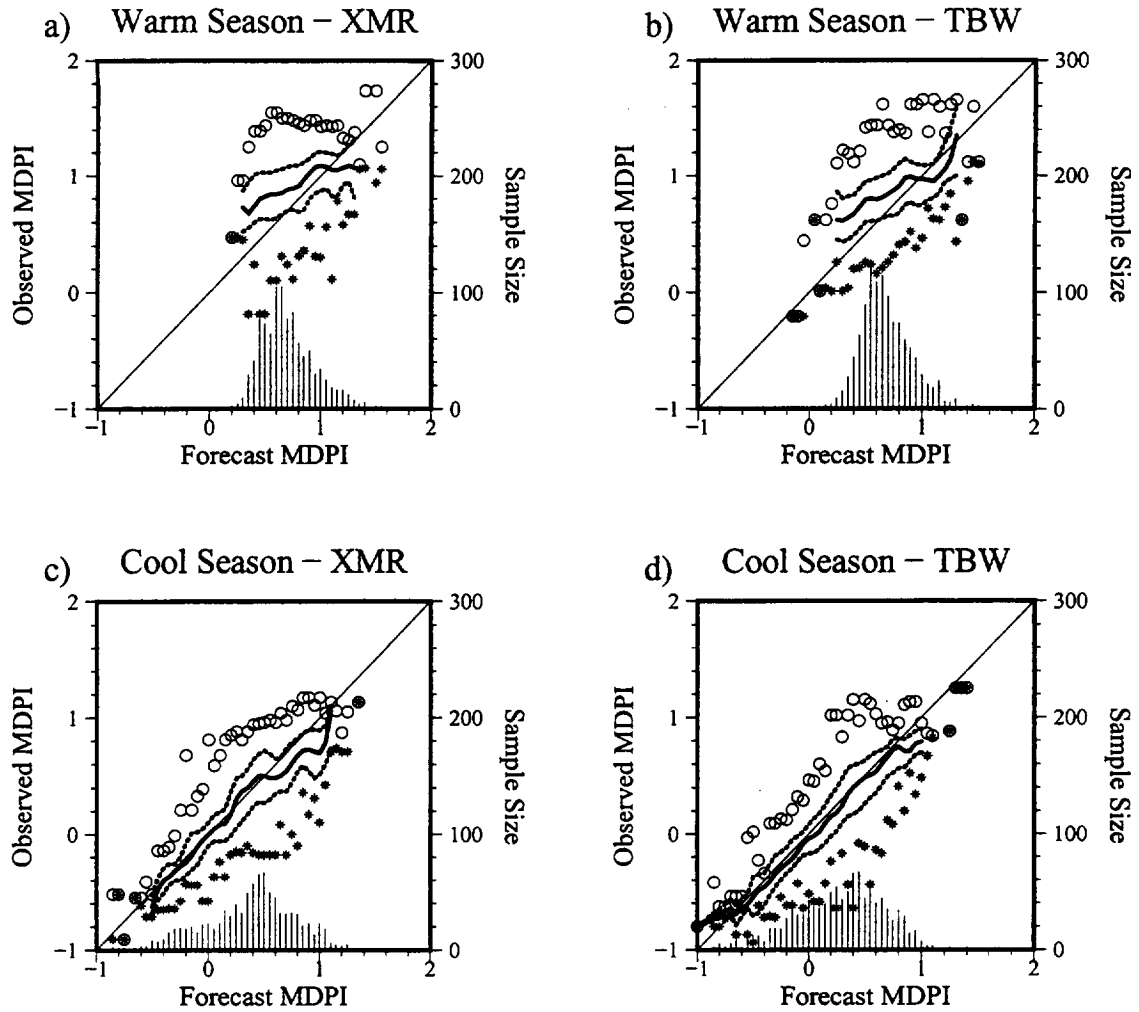


Figure 6.8. Quartile diagram showing the conditional distribution of observed MDPI associated with given forecast values and a histogram of the marginal distribution of forecasts. For each forecast of MDPI (indicated by histogram), the minimum (maximum) observations are shown by stars (circles), the .25 and .75 quartiles are shown by dotted lines, and the .50 quartile (or median) is shown by a heavy solid line. See text for additional details.

7.0 850 to 500-mb Layer-Averages

The warm and cool season bias, RMS error, and error standard deviation for 1000- to 850-mb thickness and 850- to 500-mb layer-averaged winds and relative humidity are presented for XMR, TBW, and EDW in Table 7.1. The statistics were calculated as requested by the original evaluation protocol (Table 1.1). As with convective parameters, calculations were performed using all available data from both the 0300 and 1500 UTC model runs and corresponding observations.

The RMS errors for the 1000- to 850-mb thickness are less than 12 m during both warm and cool seasons at XMR and TBW. The thickness errors are not indicated for EDW because the surface pressure at that location is often less than 1000 mb. Forecast biases for the 850- to 500-mb layer-averaged relative humidity are less than $\pm 2\%$ except at EDW during the cool season. Notably, these biases are smaller than the 2.5% observational uncertainty associated with rawinsonde observations of relative humidity (Ahnert 1991). Wind speed biases are small compared to the error standard deviation and RMS errors. These results are consistent with those shown previously in Fig. 5.4 and further indicate that the greater portion of wind speed errors are obtained from nonsystematic sources. Wind direction errors behave similarly, with near zero biases and relatively larger RMS errors and error standard deviations.

Table 7.1. Warm and cool season bias, RMS error, and error standard deviation for 1000- to 850-mb thickness and 850- to 500-mb layer-averaged relative humidity, wind speed, and wind direction.

		Bias		RMS Error		Error Std. Dev.	
		Warm	Cool	Warm	Cool	Warm	Cool
Thickness (m)	XMR	-3.02	7.27	10.31	11.77	9.86	9.26
	TBW	1.11	6.31	8.37	10.57	8.29	8.48
	EDW	—	—	—	—	—	—
Relative Humidity (%)	XMR	-.51	.81	11.63	13.62	11.62	13.60
	TBW	1.24	-1.37	12.09	12.87	12.02	12.79
	EDW	1.32	3.38	8.03	12.62	7.92	12.16
Wind Speed (m s^{-1})	XMR	.13	-.23	2.09	2.01	2.08	2.00
	TBW	-.03	-.13	1.97	2.14	1.97	2.14
	EDW	-.88	-.19	2.20	2.75	2.02	2.74
Wind Direction ($^{\circ}$)	XMR	.32	.01	42.62	22.97	42.62	22.97
	TBW	3.72	1.76	45.83	24.34	45.68	24.28
	EDW	-6.37	-4.69	38.04	38.64	37.50	38.35

8.0 Wind Regime Stratification

Within each seasonal evaluation period, the original evaluation protocol specified that the statistical verification would be stratified by average wind direction in the layer from 950 to 600 mb using seven wind regimes following Holle et al. (1992). This type of stratification was designed to determine if the model forecast errors are sensitive to specific flow regimes as defined by the layer-averaged wind directions. During the initial evaluation period in 1996, it was determined that insufficient numbers of forecast/observation pairs were available within each of the seven wind regimes. The small sample sizes made it difficult to draw meaningful conclusions regarding error characteristics as a function of flow regime (Panofsky and Brier 1958). As a simplification, westerly and easterly regimes were defined based on winds that are approximately perpendicular to Florida's east coast. In particular, the 950- to 600-mb layer-averaged winds are considered westerly between 158° and 338° and easterly outside that range. As reported by MN97, this stratification also did not reveal any substantial differences in the forecast error characteristics under easterly or westerly flow.

The extension of the evaluation for a second consecutive year helps ameliorate the difficulty of insufficient sample sizes. Again, the statistics were stratified by the 950 to 600 mb layer averaged wind direction. Examination of the stratified statistics (not shown) again did not reveal any substantial differences in forecast error from one regime to the next. On average, the observations did indicate subtle, yet expected changes based on wind direction. However, the Meso-Eta model demonstrates the capability, on average, to follow shifts in the mean observations as a function of wind regime. Therefore, the model's overall forecast error characteristics remain mostly unchanged regardless of wind regime. In most cases, the variance of the errors under the different regimes are similar. This result suggests that the day-to-day variations in the forecast errors may be more difficult to anticipate than changes in the overall error characteristics for different regimes.

9.0 Summary of Subjective Forecast Value

Statistical verification of point forecasts is a stringent measure of model performance. However, when used alone, the statistics are not enough to quantify the overall value that model guidance may add to the forecast process. This is especially true for models with enhanced spatial and temporal resolution that may be capable of generating meteorologically consistent, though not necessarily accurate mesoscale weather phenomena (e.g. Cortinas and Stensrud 1995). With this in mind, the AMU performed a subjective verification of Meso-Eta model forecasts to help quantify the added value that cannot be inferred solely from an objective evaluation. Results from the AMU's subjective verification of the Meso-Eta model over the Florida peninsula are discussed in the earlier report on the Meso-Eta evaluation (MN97) but are summarized here for completeness. Interested readers are encouraged to contact the AMU for a copy of the complete report.

The subjective evaluation included warm-season forecast exercises and phenomenological verification focusing primarily on limited case studies and seasonal evaluations of sea breezes, thunderstorms, and cold fronts. Some results from the subjective evaluation that may be important for operational forecast concerns include the following.

- Sea-breeze case studies reveal that the model generates a dynamically consistent, thermally direct circulation over the Florida peninsula, although at a larger scale than observed. Seasonal verification of sea breezes indicates that the model forecasts the occurrence of east and/or west coast sea breezes roughly 50% of the time they are observed.
- Thunderstorm verification reveals that the meso-eta model is capable of predicting areas of organized convection over portions of the Florida peninsula, particularly during the late afternoon hours. On the other hand, the model is also subject to subtle errors that can lead to incorrect forecasts of warm season convective precipitation. In particular, the model often generates excessive rainfall during the morning hours and is not capable of accurately forecasting individual thunderstorms.
- Verification of cold fronts during the cool season reveals that the model is capable of forecasting a majority of cold frontal passages through east central Florida to within ± 1 h of observed frontal passage.

Seasonal evaluations of sea breezes, thunderstorms, and cold fronts are designed to quantify how consistently the Meso-Eta model provides value in forecasts of these phenomena. Results suggest that the model forecasts over central Florida may have more value during the cool season. This statement is based on the fact that the meso-eta model resolution is not sufficient to resolve the small-scale details of sea and river/lake breeze circulations, thunderstorm outflow boundaries, and other phenomena which play a dominant role in determining the short-term evolution of weather over east central Florida during the warm season.

10.0 Summary and Lessons Learned

From May 1996 through January 1998, the AMU conducted warm- and cool-season evaluations of Meso-Eta surface and upper-air point forecast accuracy at XMR, TBW, and EDW. These three locations were selected because they are important for 45WS, NWS MLB, and SMG operational concerns. Each warm- and cool-season verification period extends from May through August and October through January, respectively. By extending the evaluation for a second consecutive year, it was possible to identify statistically significant changes in the error characteristics which developed in response to the February and August 1997 model updates (BL97; EMC 1997). The twin-season comparison of forecast accuracy is helpful for model users by highlighting the model's characteristic strengths and weaknesses before and after the incorporation of model updates. Such results are also helpful for model development efforts and emphasize the need for ongoing analysis of model errors at specific locations.

10.1 Surface Results

The error characteristics for Meso-Eta point forecasts of surface parameters are summarized in Tables 4.1-4.3. The surface error statistics are difficult to generalize because they vary widely by location, season, and time of day. The results are utilized most effectively by considering the model biases for each parameter separately and making the appropriate adjustments.

The random error component reveals substantial day-to-day variability in forecast accuracy. For many parameters, the random errors are larger than the corresponding biases, or systematic model errors. The random errors are caused primarily by the model's inability to resolve localized phenomena such as wind gusts, temperature gradients, or the effects of thunderstorms. While it is possible to partially adjust for model biases, it is much more difficult to accommodate the variability in forecast errors on any given day. It might help to compare current observations with the latest forecast guidance and make appropriate adjustments.

As modeling systems such as the Meso-Eta attain greater resolution, they tend to exhibit smaller biases and larger error standard deviations. The nonsystematic, random error component associated with any model's inability to resolve local phenomena prevents perfect forecast guidance. However, the relatively minor biases indicate that on average, point forecasts provide useful guidance for the basic sensible weather variables considered here.

Results shown in sections 4.2 and 4.3 indicate that changes to the model's physical parameterizations produced identifiable and statistically significant changes in forecast accuracy at each location. Some changes enhanced forecast accuracy while others created larger errors. It is important that model users maintain awareness of ongoing model changes. Such changes are likely to modify the basic error characteristics, particularly near the surface.

10.2 Upper-air Results

Results from the upper-air verification are summarized in Tables 5.1 and 5.2. On average, the forecast soundings at XMR and TBW during the warm season are too stable. The height of the lower tropospheric inversion at XMR and TBW was misrepresented during the cool season. Forecast biases for wind speed and direction are small at all three locations, but the random error component dominates the day-to-day variability. Given this variability, real-time assessment of forecast accuracy is necessary on any given day to help users determine if the model forecasts are consistent with current observations.

The statistics did not reveal annual changes in forecast errors that could be attributed solely to the February and August 1997 model updates. The model's systematic error growth during the forecast cycle is minimal (section 5.3), implying that all data could be combined into a single dataset regardless of duration. Moreover, since error growth is minimal, the characteristics outlined in Tables 5.1 and 5.2 apply, on average, at any time during the forecast period. This generality does not apply to the surface data where error characteristics varied with time of day.

10.3 Convective Index Results

Convective index results are summarized in section 6.1. In general, the results suggest that day-to-day fluctuations in observed convective indices at XMR and TBW are not well represented by the Meso-Eta model throughout the warm season. In addition, the conditional quartile diagrams suggest that forecasts occasionally suffer large errors in representing the true convective potential of the atmosphere. The forecast biases and variability of the errors are confirmed by the overall sample statistics (Tables 6.2 and 6.3) and support the notion that much of the error at upper levels is largely nonsystematic in nature. The convective index forecasts are most reliable overall during the cool season when, under normal circumstances, they provide little added value for most operational forecasting applications.

10.4 Layer Averages and Wind Regime Stratification

The 1000- to 850-mb thickness and 850- to 500-mb layer averaged relative humidity and winds indicate that biases in Meso-Eta forecasts for these parameters are small relative to the error standard deviations. This result suggests that the middle tropospheric errors are dominated by day-to-day variations in the difference between the forecasts and observations. Stratification of errors by the 950- to 600-mb wind direction does not reveal any substantial changes in error characteristics, and further indicates that the day-to-day variations in the forecast errors may be more difficult to anticipate than more general changes in the overall error characteristics under different wind regimes.

10.5 Lessons and Recommendations

The AMU's statistical evaluation of Meso-Eta forecast accuracy identified a few biases that may result from inadequate parameterization of physical processes near the surface. Since the model bias or systematic error generally is small, most of the total model error results from day-to-day variability in the forecasts and/or observations. To some extent, these nonsystematic errors reflect the variability in point observations that sample spatial and temporal scales of atmospheric phenomena which cannot be resolved by the model. On average, Meso-Eta point forecasts may provide useful guidance for predicting the evolution of the larger scale environment. A more substantial challenge facing model users in real time is the discrimination of nonsystematic errors that could inflate the total forecast error.

It is important that forecasters maintain an ongoing awareness of model updates and the effects that such changes will have on point forecast accuracy within their area of responsibility. While model updates are generally well tested and designed to improve forecast accuracy, the results shown here demonstrate that the desired effects do not always yield the expected improvements at every location. This may become particularly true as national-scale, operational models with greater horizontal and vertical resolutions are able to forecast explicitly the complex processes that occur within the PBL. In recent years, information documenting model updates has been made available regularly on the internet. Indeed, much of the information needed for writing this report and maintaining an understanding the model changes was obtained from an internet FAQ written expressly for this purpose (EMC 1997). As forecasters discover localized model deficiencies through ongoing real-time statistical verification strategies, results should be documented regularly and shared with model developers. As expressed by Manning and Davis (1997), "These statistics would provide additional information to model users and alert model developers to those research areas that need more attention". The additional and complementary need for subjective verification strategies in mesoscale models is discussed in the earlier report by MN97.

On 9 February 1998, NCEP upgraded the horizontal resolution of the "early" Eta model from 48 to 32 km, with an increase in vertical resolution from 38 to 45 levels (Rogers et al. 1997). In addition, a three-dimensional variational analysis scheme was implemented along with the use of a "partial" continuous Eta Data Assimilation Cycle. Aside from the differences in data assimilation methods at the time of this writing, this version of the "early" Eta is similar in resolution and dynamics to the Meso-Eta model version evaluated here. Therefore the objective verification results presented here for XMR, TBW and EDW and the subjective verification results presented in MN97 should establish a reasonable benchmark from which model users and developers may pursue ongoing Eta model verification strategies in the future.

11.0 References

- Ahnert, P. R., 1991: Precision and compatibility of National Weather Service upper air measurements. Preprints, *Seventh Symp. on Meteorological Observations and Instrumentation*, New Orleans, LA, Amer. Meteor. Soc., 221-226.
- Baldwin, M. E., and S. D. Hrebenach, 1998: Experiments with bias-corrected temperature guidance using NCEP's mesoscale Eta model. Preprints, *16th Conf. on Weather Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc., 388-389.
- Betts, A. K., F. Chen, K. E. Mitchell, and Z. I. Janjic, 1997: Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta model using FIFE data. *Mon. Wea. Rev.*, **125**, 2896-2916.
- Black, T. L., 1994: The new NMC mesoscale Eta model: description and forecast examples. *Wea. Forecasting*, **9**, 265-278.
- , M. Baldwin, K. Brill, F. Chen, G. DiMego, Z. Janjic, G. Manikin, F. Mesinger, K. Mitchell, E. Rogers, and Q. Zhao, 1997: Changes to the Eta forecast systems. NWS Technical Procedures Bulletin 441, NOAA/NWS, 31 pp. [Available from National Weather Service, Office of Meteorology, 1325 East-West Highway, Silver Spring, MD 20910.]
- Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting*, **4**, 401-412.
- Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q. Duan, and A. Betts, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE results. *J. Geophys. Res.*, **101**, 7251-7268.
- Cortinas, J. V., Jr., and D. J. Stensrud, 1995: The importance of understanding mesoscale model parameterization schemes for weather forecasting. *Wea. Forecasting*, **10**, 716-740.
- Davies-Jones, R., D. Burgess, and M. Foster, 1990: Test of helicity as a tornado forecast parameter. Preprints, *16th Conf. on Severe Local Storms*, Kananaskis Park, Alberta, Canada, Amer. Meteor. Soc., 588-592.
- desJardins, M. L., S. Jacobs, D. Plummer, and S. Schotz, 1997: N-AWIPS: AWIPS at the National Centers for Environmental Prediction. Preprints, *13th International Conf. on Interactive Information and Processing Systems*, Long Beach, CA, Amer. Meteor. Soc., 296-298.
- EMC, cited 1997: Environmental Modeling Center: Mesoscale Modeling Branch FAQ. [Available on-line from <http://nic.fb4.noaa.gov:8000/research/FAQ-eta.html>.]
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203-1211.
- Hoehne, W. E., 1980: *Precision of National Weather Service Upper Air Measurements*. NOAA Tech. Memo. NWS T&ED-16.
- Holle, R. L., A. I. Watson, R. E. Lopez, K. W. Howard, R. Ortiz, and L. Li, 1992: *Meteorological studies to improve short-range forecasting of lightning/thunderstorms within the Kennedy Space Center area*. Final Report for Memorandum of Agreement Between the Office of Space Flight, NASA and the National Severe Storms Laboratory, NOAA, Boulder, CO, 91 pp.
- Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting*, **10**, 689-707.

- Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.
- , 1996a: The Mellor-Yamada level 2.5 turbulence closure scheme in the NCEP Eta model. Research Activities in Atmospheric and Oceanic Modeling, WMO, Geneva, CAS/JSC WGNE, 4.14-4.15. [Available from World Meteorological Organization, Case Postale 2300, CH-1211 Geneva, Switzerland.]
- , 1996b: The surface layer parameterization in the NCEP Eta Model. Research Activities in Atmospheric and Oceanic Modeling, WMO, Geneva, CAS/JSC WGNE, 4.16-4.17. [Available from World Meteorological Organization, Case Postale 2300, CH-1211 Geneva, Switzerland.]
- , 1996c: The surface layer in the NCEP Eta model. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 354-355.
- Lilly, D. K. 1986: The structure, energetics and propagation of rotating convective storms. Part II: Helicity and storm stabilization. *J. Atmos. Sci.*, **43**, 126-140.
- Manning, K. W. and C. A. Davis, 1997: Verification and sensitivity experiments for the WISP95 MM5 forecasts. *Wea. Forecasting*, **12**, 719-735.
- Manobianco, J. M., and P. A. Nutter, 1997: *Evaluation of the 29-km Eta Model for Weather Support to the United States Space Program*. NASA Contractor Report CR-205409, Kennedy Space Center, FL, 91 pp. [Available from ENSCO, Inc., 1980 N. Atlantic Ave., Cocoa Beach, FL 32931.]
- McPherson, R. D., 1994: The National Centers for Environmental Prediction: Operational climate, ocean, and weather prediction for the 21st century. *Bull. Amer. Meteor. Soc.*, **75**, 363-373.
- Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc.*, **11**, 2637-2649.
- Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417-2424.
- , B. G. Brown, and Y. Chen, 1989: Diagnostic verification of temperature forecasts. *Wea. Forecasting*, **4**, 485-501.
- Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. Pennsylvania State University, University Park, 224 pp.
- Rogers, E., D. G. Deaven, and G. J. DiMego, 1995: The regional analysis system for the operational "early" Eta model: Original 80-km configuration and recent changes. *Wea. Forecasting*, **10**, 810-825.
- , T. L. Black, D. G. Deaven, and G. J. DiMego, 1996: Changes to the operational "early" Eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting*, **11**, 391-413.
- , M. Baldwin, T. Black, K. Brill, F. Chen, G. DiMego, J. Gerrity, G. Manikin, F. Mesinger, K. Mitchell, D. Parrish, Q. Zhao, 1997: Changes to the NCEP Operational "Early" Eta Analysis / Forecast System. NWS Technical Procedures Bulletin 447, NOAA/NWS. [Available from National Weather Service, Office of Meteorology, 1325 East-West Highway, Silver Spring, MD 20910.]
- Stensrud, D. J., and J. A. Skindlov, 1996: Gridpoint predictions of high temperature from a mesoscale model. *Wea. Forecasting*, **11**, 103-110.
- Tukey, J. W. 1977: *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

- Wallace, J. M., and P. V. Hobbs, 1977: *Atmospheric Science: An Introductory Survey*. Academic Press, 467 pp.
- Walpole, R. E., and R. H. Meyers, 1989: *Probability and Statistics for Engineers and Scientists*. Macmillan, 765 pp.
- Wheeler, M. M., and W. P. Roeder, 1996: Forecasting wet microburst on the Central Florida Atlantic coast in support of the United States Space Program. Preprints, *18th Conf. on Severe Local Storms*, San Francisco, CA, Amer. Meteor. Soc., 654-658.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Zhao, Q., T. L. Black, and M. E. Baldwin, 1997: Implementation of the cloud prediction scheme in the Eta model at NCEP. *Wea. Forecasting*, 12, 697-712.

Appendix A

Eta Model Overview

The primary mesoscale modeling efforts at NCEP are focused on the development of the Eta model (Rogers et al. 1995). The original version of the Eta model with a horizontal resolution of 80 km and 38 vertical layers replaced the Limited-Area Fine Mesh model in June 1993 (Black 1994). In October 1995, NCEP increased the horizontal resolution of the operational "early" Eta model from 80 km to 48 km. At the same time, a cloud prediction scheme (Zhao et al. 1997) was implemented and initial analyses were produced using the Eta Data Assimilation System (Rogers et al. 1996). In August 1995, NCEP also began running a mesoscale version of the Eta (meso-eta) model with a horizontal resolution of 29 km and 50 vertical layers (Mesinger 1996). Following model upgrades on 31 January 1996 (Chen et al. 1996; Janjic 1996a; Janjic 1996b; Janjic 1996c; Betts et al. 1997), the "early" and Meso-Eta model configurations became identical except for resolution and data assimilation procedures. The relevant numerics and physics of the Eta model are summarized in Table A.1.

NCEP implemented two major changes to the Eta model's physical parameterizations during the AMU's objective evaluation period. On 18 February 1997, components of the soil, cloud, and radiation packages were updated in both models (Betts et al. 1997; Black et al. 1997; EMC 1997). These modifications were designed to help control excessive net shortwave radiation at the ground that led indirectly to a bias in the diurnal range of surface temperatures, excessive mixing of the planetary boundary layer (PBL), and a negative bias in surface dew point temperatures. On 19 August 1997, calculation of the model's PBL depth was adjusted to correct for an underestimation of vertical moisture transport out of the lowest model layers (EMC 1997). A portion of the results shown below indicate that combined effects of these changes led to identifiable and statistically significant changes in forecast accuracy for a few selected parameters.

Table A.1. Eta model attributes from Black (1994), Janjic (1994), and Rogers et al. (1996).

Dynamics
Model top = 25 mb
Time step = 72 s
Semi-staggered Arakawa E-grid
Gravity wave coupling scheme
Silhouette-mean orography
Split-explicit time differencing
Physics
Explicit grid-scale cloud and precipitation
Modified Betts-Miller convective adjustment
Mellor-Yamada (2.5) for free atmosphere vertical turbulent exchange
Mellor-Yamada (2.0) near ground
Geophysical Fluid Dynamics Laboratory radiation scheme
Viscous sublayer over water

Appendix B

Statistical Verification Techniques

The statistical measures used here to quantify model forecast errors are the bias, root mean square (RMS) error, and standard deviation. If F represents any of the parameters under consideration for a given time and vertical level, then forecast error is defined as $\Phi' = \Phi_f - \Phi_o$ where the subscripts f and o denote forecast and observed quantities, respectively. Given N valid pairs of forecasts and observations, the bias is computed as

$$\overline{\Phi'} = \frac{1}{N} \sum_{i=1}^N \Phi'_i \quad (\text{B.1}),$$

the RMS error is computed as

$$\text{RMSE} = [\text{MSE}]^{1/2} = \left[\frac{1}{N} \sum_{i=1}^N (\Phi'_i)^2 \right]^{1/2} \quad (\text{B.2}),$$

and the standard deviation of the errors is computed as

$$\sigma' = \left[\frac{1}{N} \sum_{i=1}^N (\Phi'_i - \overline{\Phi'})^2 \right]^{1/2} \quad (\text{B.3}).$$

In equation (B.3), N is used rather than $N-1$ so that a decomposition following Murphy (1988, Eq. 9) could be applied to the MSE:

$$\text{MSE} = \overline{\Phi'}^2 + \sigma'^2 \quad (\text{B.4}).$$

Therefore, the total model error consists of contributions from model biases ($\overline{\Phi'}^2$) and random variations in the forecast and/or observed data (σ'^2). Note that if the model bias or systematic error is small, most of the MSE is due to random, nonsystematic type variability in the errors. Murphy's (1988) decomposition of the MSE considered individually the error contributions from the model bias and from the sample variances and covariance of the forecasts and observations. Here, Eq. B.4 represents an algebraic simplification of that decomposition and quantifies the portion of the MSE that is due to the bias and the variance of the forecast errors.

Tests are applied to the surface data in order to determine if model updates led to statistically significant changes in mean forecast error between the 1996 and 1997 warm and cool season periods. Following the Central Limit Theorem as described in most statistical texts, it is assumed that the sampling distribution for the difference in mean forecast error between 1996 and 1997 is approximately normal. Sample sizes of $O(100)$ for each season enable use of the standardized Z statistic where

$$Z = \frac{\overline{\Phi'}_{97} - \overline{\Phi'}_{96}}{\left\{ \delta_{96} \left[(\sigma'_{96})^2 / N_{96} \right] + \delta_{97} \left[(\sigma'_{97})^2 / N_{97} \right] \right\}^{1/2}} \quad (\text{B.5}),$$

the variance inflation factor, $\delta = (1 + r)/(1 - r)$, and r is the lag-1 day autocorrelation for each seasonal time series of data. The variance inflation factor helps prevent the overestimation of Z by adjusting the variance of the sampling distribution to account for the influence of serial dependence, or day-to-day persistence, within the seasonal time series average (Wilks 1995). A two-tailed comparison of Z to the normal distribution using a 99% confidence level has critical values of ± 2.58 (Walpole and Meyers 1989). Calculated values of Z that lie outside this critical range indicate that the data are able to support a statistically significant difference between the 1996 and 1997 seasonal mean forecast errors.

The statistical significance of upper-level systematic error growth from early to latter stages of the forecast cycle is determined using a paired Z statistic. The paired Z statistic normalizes the seasonally averaged difference in forecast error between two times during the i th cycle by the associated sample standard deviation. The covariance between errors in the early and latter stages of the forecast is included because the parameters from the i th cycle are not independent and do not necessarily have equal variances (Walpole and Meyers 1989). Here, the paired Z statistic is denoted by Z' where

$$Z' = \frac{\sum_{i=1}^N (\Phi'_{2i} - \Phi'_{1i})}{\left\{ N [(\sigma'_1)^2 + (\sigma'_2)^2 - (\sigma'_{12})^2] \right\}^{1/2}} \quad (\text{B.6}).$$

The subscripts 1 and 2 denote variables from the i th forecast cycle verifying at 6-9 h and 30-33 h, respectively. The times used for verification are separated by 24 h and are taken at forecast durations that vary slightly according to balloon release times. Other notations are as above except that $(\sigma'_{12})^2$ denotes the sample covariance. Again using a 99% confidence level, values of Z' that lie outside the critical values of ± 2.58 indicate that the data are able to support a statistically significant 24-h systematic error growth in the upper-air forecasts.

Appendix C

Upper Level Wind Errors Listed as a Function of Height

As specified in the original evaluation protocol, objective verification results are listed as a function of geopotential height. The wind variables are of primary interest for applications to Shuttle flight rules and are listed in tabular format for direct reference. Plots of wind speed and direction errors as a function of height (not shown) exhibit exactly the same characteristics as those shown in Figs. 5.3 and 5.4. Therefore, the discussions in sections 5.3 and 5.4 are applicable to the wind speed and direction errors listed below. Although results for u and v wind components are not discussed in previous sections they are generally consistent with results for wind speed. As discussed at the beginning of section 5, substantial differences do not appear to exist between the 0300 and 1500 UTC forecast cycles. Moreover, the statistics do not reveal any obvious differences when the results are stratified by the 950- to 600-mb layer-averaged wind direction. For these reasons, all available data from each season are included in the statistical calculations below. Results from TBW are not shown but are in fact very similar to those shown below for XMR. At EDW, results below 3000 ft are not available since the lowest level of data extracted from model point forecasts at that location is at an elevation of 3238 ft.

Table C.1. Warm season biases for wind variables at XMR and EDW as a function of geopotential height.

Height (kft):	U-wind (m s^{-1})		V-wind (m s^{-1})		Wind Speed (m s^{-1})		Wind Direction ($^{\circ}$)	
	XMR	EDW	XMR	EDW	XMR	EDW	XMR	EDW
1	.21	---	-.27	---	.16	---	1.37	---
2	-.14	---	.13	---	.48	---	1.31	---
3	-.44	---	.41	---	.47	---	3.05	---
5	-.64	-.80	.65	-.92	.33	-1.24	-6.45	-2.62
10	-.02	-.15	-.40	.27	.07	-1.04	-1.04	-6.91
12	.18	-.04	-.60	.26	-.01	-.89	2.17	-7.53
20	-.52	-.71	-.20	1.00	-.34	.05	-.72	-9.22
25	-.46	-.98	-.37	1.25	-.22	-.44	-1.45	-10.29
28	-.11	-.62	-.39	1.53	-.13	-.24	1.14	-8.59
35	.53	-.42	-.30	1.29	.16	-.15	-1.25	-5.22
38	.68	-.38	.04	1.24	-.01	.02	-2.65	-4.05
45	.90	-.75	.68	1.35	.48	-.79	-4.85	-2.81
50	1.20	-.90	.28	.40	.85	-1.05	-4.89	-1.91
55	1.70	1.09	.34	.71	.34	.01	-3.21	-.39
60	3.50	2.89	1.35	-.65	-.72	-.23	3.27	10.64
70	8.15	5.04	4.29	-.59	-3.66	-1.70	35.63	9.87

Table C.2. Cool season biases for wind variables at XMR and EDW as a function of geopotential height.

Height (kft):	U-wind (m s^{-1})		V-wind (m s^{-1})		Wind Speed (m s^{-1})		Wind Direction ($^{\circ}$)	
	XMR	EDW	XMR	EDW	XMR	EDW	XMR	EDW
1	-.16	---	-.28	---	.34	---	-.23	---
2	-.67	---	-.12	---	.86	---	.23	---
3	-.80	---	.25	---	.53	---	3.41	---
5	-.65	-2.07	.46	.05	-.03	-.93	.32	-.20
10	-.33	-1.03	-.07	1.15	-.22	-.45	-.10	-4.82
12	-.41	-.79	-.18	1.74	-.36	-.59	.44	-2.29
20	-.71	.95	-.34	2.11	-.66	.75	1.21	-7.02
25	-.37	1.18	-.43	2.44	-.48	.68	1.76	-8.70
28	-.43	1.75	-.25	2.57	-.54	1.87	1.21	-7.29
35	-.10	2.27	.06	2.73	-.39	1.43	-.17	-7.35
38	-.18	2.18	.38	2.71	-.32	1.61	-1.34	-6.48
45	.70	2.96	.63	1.01	.84	2.31	-1.10	-3.88
50	.45	1.80	-.05	2.42	.27	1.23	-.12	-7.11
55	1.75	2.09	-.34	.33	1.89	1.85	-1.03	-5.79
60	5.96	2.83	-.02	.38	5.16	2.76	-5.16	-4.43
70	9.47	4.51	1.54	-1.94	7.14	4.47	-.12	4.69

Table C.3. Warm season RMS errors for wind variables at XMR and EDW as a function of geopotential height.

Height (kft):	U-wind (m s^{-1})		V-wind (m s^{-1})		Wind Speed (m s^{-1})		Wind Direction ($^{\circ}$)	
	XMR	EDW	XMR	EDW	XMR	EDW	XMR	EDW
1	3.01	---	3.10	---	2.66	---	53.20	---
2	2.87	---	2.90	---	2.74	---	51.81	---
3	2.85	---	2.93	---	2.80	---	54.56	---
5	3.05	3.95	3.18	3.14	2.93	3.78	57.16	47.05
10	3.01	3.26	2.85	3.01	2.90	2.93	49.62	48.27
12	3.00	3.38	2.89	3.25	2.99	3.32	48.97	43.23
20	3.09	3.70	2.86	3.65	2.75	3.48	45.19	34.39
25	3.42	3.87	3.12	3.77	3.01	3.62	44.90	27.68
28	3.81	3.49	3.33	4.17	3.55	3.52	44.00	26.99
35	4.99	4.33	4.62	4.46	4.78	4.47	42.91	23.54
38	5.57	4.52	5.21	4.83	5.33	4.71	40.93	19.14
45	5.36	4.13	5.41	4.92	5.26	4.50	37.65	20.37
50	4.80	3.77	4.40	4.05	4.67	3.98	42.33	24.77
55	4.37	4.07	3.85	4.09	3.89	3.89	52.14	37.31
60	4.72	4.88	3.56	3.58	3.94	3.60	57.40	55.59
70	8.88	6.05	5.69	3.22	6.17	4.94	82.92	59.58

Table C.4. Cool season RMS errors for wind variables at XMR and EDW a function of geopotential height

Height (kft):	U-wind (m s^{-1})		V-wind (m s^{-1})		Wind Speed (m s^{-1})		Wind Direction ($^{\circ}$)	
	XMR	EDW	XMR	EDW	XMR	EDW	XMR	EDW
1	2.67	---	3.04	---	2.64	---	38.11	---
2	2.66	---	3.16	---	2.84	---	38.52	---
3	3.01	---	3.13	---	3.04	---	40.01	---
5	3.33	4.38	3.37	3.23	3.27	3.68	40.13	61.70
10	2.90	4.30	3.26	4.53	2.89	3.81	32.33	46.99
12	2.95	4.87	3.18	4.96	2.92	4.22	30.83	44.16
20	3.64	5.50	3.61	6.98	3.62	5.72	23.51	31.21
25	3.76	6.99	3.97	7.89	3.75	6.90	20.17	31.57
28	4.08	7.10	4.09	8.04	4.13	7.41	16.09	26.88
35	5.57	7.99	5.26	10.33	5.56	8.05	12.61	23.51
38	5.16	7.94	5.06	9.90	5.14	8.81	12.76	24.23
45	5.11	8.24	5.71	7.21	5.23	7.51	11.41	21.71
50	5.22	6.52	5.17	6.60	5.11	5.99	11.82	23.85
55	4.53	5.70	4.33	6.84	4.46	5.73	16.79	27.62
60	7.08	6.08	3.88	4.91	6.50	6.30	37.16	22.88
70	10.15	6.95	4.57	5.04	8.81	6.89	68.87	33.26

Table C.5. Warm season error standard deviations for wind variables at XMR and EDW as a function of geopotential height.

Height (kft):	U-wind (m s^{-1})		V-wind (m s^{-1})		Wind Speed (m s^{-1})		Wind Direction ($^{\circ}$)	
	XMR	EDW	XMR	EDW	XMR	EDW	XMR	EDW
1	3.00	---	3.09	---	2.66	---	53.18	---
2	2.87	---	2.90	---	2.70	---	51.79	---
3	2.82	---	2.90	---	2.76	---	54.48	---
5	2.98	3.87	3.12	3.01	2.91	3.57	56.80	46.98
10	3.01	3.26	2.83	3.00	2.90	2.74	49.61	47.78
12	2.99	3.38	2.83	3.24	2.99	3.20	48.92	42.57
20	3.05	3.63	2.86	3.51	2.73	3.48	45.18	33.13
25	3.39	3.74	3.10	3.56	3.00	3.59	44.88	25.69
28	3.81	3.43	3.31	3.88	3.55	3.51	43.99	25.59
35	4.96	4.31	4.61	4.26	4.77	4.47	42.89	22.96
38	5.53	4.50	5.21	4.67	5.33	4.71	40.84	18.70
45	5.29	4.06	5.37	4.73	5.24	4.43	37.33	20.18
50	4.65	3.66	4.39	4.03	4.60	3.84	42.05	24.69
55	4.02	3.92	3.83	4.02	3.88	3.89	52.04	37.31
60	3.17	3.93	3.30	3.52	3.87	3.60	57.31	54.56
70	3.54	3.34	3.74	3.17	4.96	4.63	74.87	58.76

Table C.6 Cool season error standard deviations for wind variables at EDW as a function of geopotential height.

Height (kft):	U-wind (m s^{-1})		V-wind (m s^{-1})		Wind Speed (m s^{-1})		Wind Direction ($^{\circ}$)	
	XMR	EDW	XMR	EDW	XMR	EDW	XMR	EDW
1	2.66	---	3.03	---	2.62	---	38.11	---
2	2.57	---	3.16	---	2.70	---	38.52	---
3	2.91	---	3.12	---	2.99	---	39.86	---
5	3.26	3.86	3.33	3.23	3.27	3.56	40.13	61.70
10	2.88	4.17	3.25	4.39	2.88	3.79	32.33	46.74
12	2.92	4.81	3.17	4.65	2.89	4.18	30.83	44.10
20	3.57	5.42	3.59	6.65	3.56	5.67	23.48	30.41
25	3.74	6.89	3.95	7.50	3.72	6.87	20.09	30.35
28	4.06	6.88	4.08	7.62	4.09	7.17	16.05	25.88
35	5.57	7.66	5.26	9.96	5.55	7.92	12.61	22.34
38	5.16	7.64	5.04	9.52	5.13	8.67	12.69	23.34
45	5.06	7.69	5.68	7.14	5.17	7.15	11.36	21.36
50	5.20	6.27	5.17	6.14	5.10	5.86	11.82	22.77
55	4.18	5.30	4.32	6.83	4.04	5.43	16.76	27.01
60	3.81	5.38	3.88	4.90	3.95	5.66	36.80	22.45
70	3.65	5.29	4.30	4.65	5.16	5.25	68.87	32.92

Appendix D

Conditional Quartile Diagrams

Although traditional statistics such as the bias and RMS error are useful for quantifying overall forecast accuracy within a particular sample of data, they fail to describe completely the character of the relationship between given forecasts and their corresponding observations. Murphy et al. (1989) demonstrate that such relationships are revealed, in part, by considering a factorization of the joint distribution between forecasts (f) and their observations (x). Given a set of data containing f and x , the joint distribution $p(f,x)$ specifies the relative frequency of occurrence for particular combinations of f and x . The factorization for the joint distribution considered here is $p(f,x) = p(x|f)p(f)$, where $p(f)$ is the marginal distribution of the forecasts and $p(x|f)$ is the conditional distribution of the observations given the forecasts. Given the set of available forecasts, f , the marginal distribution $p(f)$ describes the relative frequency of occurrence for each forecast value (or discretized range of values). The conditional distribution of the observations given the forecasts, $p(x|f)$, specifies the marginal distribution of the subset of all observations x which correspond to a particular forecast.

The factorization of $p(f,x)$ is useful because $p(x|f)$ and $p(f)$ specify the calibration and refinement of the forecasts, respectively (Murphy et al. 1989). A set of forecasts are perfectly calibrated if, for each forecast f , the mean observation is equal to f . A set of forecasts are completely unrefined if the same forecast value is produced on each occasion. Therefore, a set of forecasts may be considered well refined and reliable if the marginal distribution of forecasts $p(f)$ covers an appropriate range of values and the average of the conditional distribution of observations is equal to the forecast, f .

The relationship between $p(x|f)$ and $p(f)$ may be interpreted graphically using a conditional quartile diagram (Murphy et al. 1989). For each set of convective index forecasts, the marginal distribution of the forecast values, $p(f)$, is displayed (Figs. 6.1 - 6.8) using histograms. In addition, the conditional distribution of the observations, $p(x|f)$, is summarized for each given forecast value by the median, interquartile range (IQR), and minimum and maximum values. The median represents the middle value of a set of ordered data and is alternately called the 50th percentile or 2nd quartile. Similarly, the IQR represents the difference between the 3rd and 1st quartiles (75th and 25th percentiles). The quartile lines extend across those index values which are forecast on at least five occasions and have been smoothed using a simple 3-point hanning algorithm (Tukey 1977). Deviations of the conditional medians from the 45° reference line reveal that the forecasts are conditionally biased. Specifically, deviations above (below) this line indicate forecasts from a particular category which are more often smaller (larger) than observed. More generally, the diagrams utilize five data points (minimum, 1st, 2nd and 3rd quartiles, and maximum) to describe the character of the conditional distributions for the observations associated with each given forecast value.

Although the terminology is complex, the conditional quartile plots aid operational users by serving as a kind of historical lookup diagram which helps visualize the relationship between given forecasts and observations. For example, if the Meso-Eta model provides a precipitable water (PWAT) forecast of 30 mm at XMR during the warm season, Fig. 6.1a indicates that such a forecast is traditionally uncommon and conditionally unbiased, and that 50% of the observations (i.e., the IQR) were within about ± 5 mm of that given forecast value.

Appendix E

Summary of Meso-Eta Model Point Forecast Error Characteristics

for

XMR; TBW; EDW

1. Surface Parameter Forecasts (Tables E.1-E.3)

Error characteristics for surface parameter forecasts vary widely by location, season, and time of day. The statistics can be utilized most effectively by considering the model biases for each parameter separately. For example, the fact that Meso-Eta wind speed forecasts are too fast on average at XMR (Table 1) suggests that forecast accuracy might be improved by adjusting such guidance to lower speeds. Similar adjustments should be made to accommodate the biases identified for other parameters.

The random error component reveals substantial day-to-day variability in forecast accuracy. The random errors are caused primarily by the model's inability to resolve localized phenomena such as wind gusts, temperature gradients, or the effects of thunderstorms. While it is possible to partially adjust for model biases, it is much more difficult to accommodate the variability in forecast errors on any given day. It might help to compare current observations with the latest forecast guidance and make appropriate adjustments.

On average, the model provides useful guidance for time-averaged environmental parameters such as METAR observations. However, the model does not have sufficient resolution to forecast events such as peak wind gusts.

It is important that users maintain awareness of ongoing model changes. Such changes are likely to modify the basic error characteristics, particularly near the surface.

Table E.1. Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at XMR during the warm (May through Aug 1997) and cool (Oct 1997 through Jan 1998) seasons. A range of errors reveals fluctuations with time of day as demonstrated in sections 4.2 and 4.3.

Variable	Season	RMS	Bias	Std Dev	Interpretation
Sea-level Pressure (mb)	Warm	1	–1 to 0	1	Forecasts tend to be slightly lower than observed.
	Cool	1	±0.5	1	Small, variable forecast bias with random errors of 1 mb.
Temp. (°C)	Warm	1 to 2	±1	1 to 2	Forecasts are slightly warm in afternoon, slightly cool at night. Large random error component.
	Cool	2	0 to 1	2	Slight warm bias throughout the forecast cycle. Random error contributes more than bias.
Dew Point (°C)	Warm	1 to 2	–1 to 0	1 to 2	Forecasts are slightly dry on average. Random error contributes more than bias.
	Cool	1 to 3	0 to 2	1 to 2	Forecasts are typically wetter than observed.
Wind Speed (m s ^{–1})	Warm	2	0 to 2	1 to 2	Forecast winds are too fast on average.
	Cool	2 to 3	1 to 3	1.5	Forecast winds are too fast on average.
Wind Dir. (°)	Warm	50 to 70	±10	50 to 70	Forecasts are nearly unbiased although random errors are large.
	Cool	40 to 60	±10	40 to 60	Same as warm season except random errors are slightly smaller.

Table E.2. Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at TBW during the warm (May through Aug 1997) and cool (Oct 1997 through Jan 1998) seasons. A range of errors reveals fluctuations with time of day as demonstrated in sections 4.2 and 4.3.

Variable	Season	RMS	Bias	Std Dev	Interpretation
Sea-level Pressure (mb)	Warm	1	–1 to 0	1	Forecasts tend to be slightly lower than observed.
	Cool	1	±0.5	1	Small, variable forecast bias with random errors of 1 mb.
Temp. (°C)	Warm	2.5	–3 to 1	1 to 2	Forecasts are too warm in the afternoon, too cool at night.
	Cool	1 to 3	–1 to 3	1 to 2	Forecasts are too warm in the afternoon, too cool at night.
Dew Point (°C)	Warm	1 to 2	–1 to 0	1 to 2	Forecasts are slightly dry on average. Random error contributes more than bias.
	Cool	1 to 3	0	1 to 3	Forecasts are unbiased but random errors reduce accuracy.
Wind Speed (m s ^{–1})	Warm	1.5	±1	1 to 2	Small forecast bias. Random error contributes more than bias.
	Cool	2	0 to 1	1.5	Forecast winds are slightly fast on average.
Wind Dir. (°)	Warm	50 to 80	–30 to 0	50 to 80	Forecast winds should be backed slightly to better match the observations.
	Cool	30 to 50	–20 to 0	30 to 50	Same as warm season except random errors are smaller.

Table E.3. Summary of Meso-Eta forecast biases (forecast – observed), RMS errors, and error standard deviations for surface parameters at EDW during the warm (May through Aug 1997) and cool (Oct 1997 through Jan 1998) seasons. A range of errors reveals fluctuations with time of day as demonstrated in sections 4.2 and 4.3.

Variable	Season	RMS	Bias	Std Dev	Interpretation
Sea-level Pressure (mb)	Warm	1 to 3	–2 to 0	1.5	Forecasts tend to be lower than observed.
	Cool	2 to 3	0 to 3	2	Forecasts tend to be greater than observed.
Temp. (°C)	Warm	3 to 6	–6 to –2	1 to 3	Forecasts are too cold on average.
	Cool	3 to 5	–4 to 0	2 to 4	Forecasts are too cold on average, especially during the daytime.
Dew Point (°C)	Warm	3 to 9	0 to 8	3 to 5	Forecasts are too moist on average, especially during the daytime.
	Cool	3 to 6	–1 to 5	3.5	Forecasts are mostly wetter than observed, especially during the daytime.
Wind Speed (m s ^{–1})	Warm	2 to 6	–7 to –1	1.5 to 3	Forecasts too slow on average, especially during the daytime.
	Cool	2 to 3	–2 to 0	2	Forecasts too slow on average.
Wind Dir. (°)	Warm	20 to 90	0 to 30	20 to 90	Forecast winds should be veered slightly overnight to better match the observations.
	Cool	60 to 90	0 to 30	60 to 90	Same as warm season.

2. Upper-Air Forecasts (Tables E.4-E.5)

Upper-air statistics were computed using all available data collected during the twin warm and cool season periods. Since model biases do not significantly increase with time at upper-levels, the error characteristics outlined in Tables E.4 and E.5 apply at any time during the forecast period. This generality does not apply to surface forecasts where error characteristics vary with time of day.

Table E.4. Summary of Meso-Eta upper-air forecast error characteristics at XMR and TBW.

Warm Season (May – Aug)	Cool Season (Oct – Jan)
On average, forecasts are about 1 °C too cold below 700 mb and 1 to 2 °C too warm above 700 mb.	The height of the lower tropospheric temperature inversion is often overforecast, thereby creating a 2 °C cold bias near the 700-mb level.
Forecasts are too dry below 800 mb and too moist above 500 mb.	Wind speed forecasts are about 1 m s ⁻¹ too slow in the middle troposphere and about 1 m s ⁻¹ too fast in the upper troposphere.
The temperature and moisture biases indicate that forecast soundings are too stable on average. This could be a consequence of the model's convective rainfall parameterization.	Wind direction forecast biases are less than ±10°, but the random error component of 10 to 40° dominates the day-to-day variability.
Wind speed forecasts are nearly unbiased in the lower and middle troposphere, but are typically too fast above 400 mb.	
Wind direction forecast biases are less than ±10° but the random error component of 40 to 60° dominates the day-to-day variability.	

Table E.5. Summary of Meso-Eta upper-air forecast error characteristics at EDW.

Warm Season (May – Aug)	Cool Season (Oct – Jan)
Temperature biases are less than ±1 °C.	A strong cold bias exists in the forecasts below 700 mb. The bias exceeds –4 °C near the surface.
Forecasts tend to retain greater amounts of moisture than observed except near the 600 mb level.	Forecasts are too moist near the surface, and too dry above 800 mb.
Wind speed forecasts are 1 to 2 m s ⁻¹ too slow, but the random error component of 3 to 5 m s ⁻¹ dominates the day-to-day variability.	Wind speeds are 1 to 2 m s ⁻¹ too slow on average except near the tropopause. The random error component exceeds 6 m s ⁻¹ .
On average, wind direction forecasts are backed about 10° relative to observations. The random error component of 30 to 60° dominates the day-to-day variability.	On average, wind direction forecasts are backed about 10° relative to observations. The random error component of 30 to 90° dominates the day-to-day variability.

3. Convective Index Forecasts (XMR; Table E.6)

Convective indices are vertically integrated quantities. Therefore, slight errors in the forecast vertical profiles of temperature, moisture, and/or winds are exacerbated through the computation of convective index error characteristics. The day-to-day variability in observed convective indices is not predicted well by the Meso-Eta model. The errors tend to smallest during the cool season when, under normal circumstances, they provide little added value for operational forecasting purposes.

Table E.6. Meso-Eta point forecast error characteristics for convective indices at XMR during the 1996 and 1997 warm seasons (May-Aug).

Index	Forecast Error Characteristics
Precipitable Water	Forecasts have a slight dry bias, but are generally accurate across a wide range of values.
Lifted Index	Forecasts are more stable than observed. The forecasts tend to be most accurate when their values are around -3 to -4 °C but the day-to-day variations are not handled well.
K-index	Forecast biases are small but a large random error component limits their utility.
LCL	Forecast accuracy decreases at lower pressures (greater heights).
CAPE	Forecasts are too small (stable) on average and are susceptible to very large errors.
0-3 km Helicity	Forecasts tend to overestimate the magnitude of the vertically integrated wind shear.
MDPI	Forecasts are most reliable when values are near 1.0 but a large random error component limits their utility.

NOTICE

Mention of a copyrighted, trademarked or proprietary product, service, or document does not constitute endorsement thereof by the author, ENSCO, Inc., the AMU, the National Aeronautics and Space Administration, or the United States Government. Any such mention is solely for the purpose of fully informing the reader of the resources used to conduct the work reported herein.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE July 1998	3. REPORT TYPE AND DATES COVERED Contractor Report	
4. TITLE AND SUBTITLE An Extended Objective Evaluation of the 29-km Eta Model for Weather Support to the United States Space Program			5. FUNDING NUMBERS C-NAS10-96018	
6. AUTHOR(S) Paul Nutter, John Manobianco				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ENSCO, Inc., 1980 N. Atlantic Ave., Cocoa Beach FL 32931			8. PERFORMING ORGANIZATION REPORT NUMBER 98-003	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) NASA, John F. Kennedy Space Center, Code PH-B3, Kennedy Space Center, FL 32899			10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA CR-1998-207910	
11. SUPPLEMENTARY NOTES Subject Cat.: #47 (Weather Forecasting)				
12A. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited			12B. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) This report describes the Applied Meteorology Unit's objective verification of the National Centers for Environmental Prediction 29-km eta model during separate warm and cool season periods from May 1996 through January 1998. The verification of surface and upper-air point forecasts was performed at three selected stations important for 45th Weather Squadron, Spaceflight Meteorology Group, and National Weather Service, Melbourne operational weather concerns. The statistical evaluation identified model biases that may result from inadequate parameterization of physical processes. Since model biases are relatively small compared to the random error component, most of the total model error results from day-to-day variability in the forecasts and/or observations. To some extent, these nonsystematic errors reflect the variability in point observations that sample spatial and temporal scales of atmospheric phenomena that cannot be resolved by the model. On average, Meso-Eta point forecasts provide useful guidance for predicting the evolution of the larger scale environment. A more substantial challenge facing model users in real time is the discrimination of nonsystematic errors that tend to inflate the total forecast error. It is important that model users maintain awareness of ongoing model changes. Such changes are likely to modify the basic error characteristics, particularly near the surface.				
14. SUBJECT TERMS Model Verification, Eta Model, Mesoscale Weather Forecasting, Numerical Weather Prediction			15. NUMBER OF PAGES 64	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT NONE	

